CONCEPTION AND CAUSATION: SELECTED EARLY PHILOSOPHICAL PAPERS

BY J.-M. KUCZYNSKI

Conception and Causation

J.-M. Kuczynski

Emotions and the Content of Moral Statements

Why conceivability sometimes entails possibility:

a Neo-Cartesian argument for dualism

Program-causes and the Inertness of the Mental

THE USELESSNESS OF THE RAMSEY-LEWIS SENTENCE

Intentionalism

Simulated versus actual thought

Davidson's Misconceptions about Rationality

Jerry Fodor's Pseudo-computationalism

Can functionalism be articulated without vicious circularity?

How to be a non-Atomist about Conception without being a Holist

The nature of emotions in relation to the central question of meta-ethics

NO SUCH THINGS AS MENTALESE

NO SUCH THING AS MENTALESE REVISITED

The Bundle Theory of Personal Identity

Conception as Knowledge of Uniquely Individuating Descriptions

Intentionalism and Foundationalism

Propositions, Logical Truth, and the Analytic-Synthetic Distinction

Causal Efficacy and the Erroneousness of Functionalism and Content-externalism

Program-causes and Mental Efficacy

Externalism and the Veil of Perception

Representationalism

Review of Erin McReady's review of Conceptual Atomism and the Computational Theory of Mind

The Computational Theory of Mind

Hempel's paradox solved

What is Formal Truth?

Time-travel: An incoherent concept

Probabilistic Causation

Propositions as Sets of Properties, Truth as Instantiatedness

Propositions as sets of properties, truth as set-member-instantiatedness (the short version)

Fodor on Concepts: How to be a non-Atomist without being a Holist

Introduction

The papers in the volume put forth clear answers to questions relating to:

*The nature of emotion
*The relevance of a correct analysis of emotion to ethics
*The nature of conception
*The nature of certain kinds of causation, and finally
*The bearing of correct theories of conception and causation on questions as to the nature of mind.

Proof that we have always known the theory of evolution to be true

These days it is pretty much agreed that human beings evolved from other, "lower" life forms and that, ultimately, we come from the same stock as monkeys, reptiles, and insects. This discovery was, of course, a massive blow to our self-esteem. Given that our forbears are single-celled organisms, it becomes hard to see how, in any cosmic sense, we could be special. We are, it would seem, just very sophisticated protozoa.

Once it is agreed that we came into existence by a process of evolution, it becomes hard to maintain that we were created by an almighty God. (Of course, it could be said that he created us by creating organisms from which we evolved. But somehow that position isn't very satisfying.) Further, the creatures from which we evolved -- reptiles, amphibians, single-celled organisms – are ones which we are used to holding in contempt. Now that we know about evolution, we cannot hold them in contempt. For we *are* them, at least in a phylogenetic sense.

There is at least one other reason why the theory of evolution upsets us. It makes it very hard to believe in an afterlife. We don't believe that amoebas or lizards or baboons go to heaven. So given that we evolved from such creatures, or from creatures of their ilk, it seems very arbitrary to maintain that *we* go to heaven when we die. Which species was the first to have an afterlife? Homo Erectus? Australopithecus? Cro-magnon? Any choice seems arbitrary. The only *non*-arbitrary thing to say is that since baboons don't go to heaven, neither do we; for we are just baboons with big brains.

It is generally thought that, prior to Darwin's work in the mid 19th century, people didn't know about evolution. Before Darwin, we believed that we *were* cosmically special and that we were of divine, not reptilian,

origin.

In any case, it is generally *held* that we believed this. I believe that people have *always* known about evolution, at least at some level. I cannot prove this here; but the idea is, I think, worth considering and developing. Also, it isn't hard to make some case for it. If we leave aside our wishes, and consider the facts in plain view, the most obvious thesis is *not* that we were of divine origin but that, on the contrary, we *are* related to dogs, monkeys, and other "lower" organisms. In fact, it is painfully obvious that we are related to such creatures. And this makes it hard to believe that people haven't *always* known that we are related to them and thus come from a common stock. Of course, once it is granted that monkeys, lizards, and amoebas are our brethren, it is a very short step (if it is a step at all) to seeing that human beings emerged by from other "lower" life-forms. So I would suggest that – in some way, at some level -- people knew that our forbears were "lower" organisms and, consequently, that people evolved from such creatures.

Of course, people didn't know about evolution at a *conscious* level. They repressed this knowledge; they knew it only unconsciously. Humankind's insistence that it was of divine origins can be seen as a way of warding off or defusing this ugly piece of unconscious knowledge.

None of this is to belittle Darwin's discoveries. Oftentimes it takes genius to make conscious what is known unconsciously. In fact, given how strong our resistances were to believing in evolution – given what a massive emotional investment people had in *not* believing in evolution - the fact that Darwin's positions have received near universal acceptance is an unambiguous indication of his genius.

Why believe that we have, at some level, *always* known about evolution? First of all, there are extremely obviously anatomical similarities between us and other, "lower" animals. We have the same basic anatomical configuration

as virtually all mammals and many reptiles and amphibians. For example, we have two eyes, four limbs, a spinal cord. And anyone who has seen the innards of a human being and those of, say, a cow or a dog knows that the differences are miniscule compared to the similarities. (Presumably people in earlier cultures were more likely to be involved preparing their own food and were more likely than people today to have direct knowledge of these similarities.)

Of course, we are very different from many animals. We are, at least in terms of overt anatomy, quite different from snakes and squids. (At the same time, there are some very obvious similarities. No one can look at a squid in a museum without having the haunting feeling that one is looking at *another* thinking being. The eyes of a squid are eerily human.) But it is very easy to establish an *indirect* resemblance between ourselves and practically *any* species. Snakes directly resemble certain four-legged lizards. (There are some lizards that have four tiny limbs, but otherwise look like snakes.) And it is easy to establish a direct resemblance between *those* reptiles and others that have four large limbs and a tail. *Those* obviously resemble various mammals, at least in terms of their overt morphology. And so on, until we arrive at a species – like the Orangutan – which obviously resembles us. So even though there are some organisms we do not obviously resemble, we *indirectly* resemble all organisms.

It does not take any acumen to see that, in terms of morphology, we are very similar to dogs and monkeys. Any two-year-old notices the similarities. (Actually, I would not be surprised if, in the earliest phases of their cognitive development, people thought humans and monkeys and dogs *all* belonged to the same species.) Now when a number of things resemble one another, it is simple common sense to suppose that they all have a common cause. If four pairs of shoes all have the same logo on them, they were probably made by

the same company. If four houses have the same architecture, they were probably designed – if not by the same person – then by people belonging to the same architectural school. And the different adherents of that school can trace their lineage back to a common source (e.g. Frank Lloyd Wright). So *ultimately* the houses in question probably had a common origin.

This is not a rarified point about scientific methodology. This is how people *actually* think. *Every* human being, when confronted with two very similar things, instinctively looks for a common cause. In fact, the ability to accept *coincidence* is the mark of *non*-instinctive thought. (A person in a reduced psychological condition – e.g. schizophrenia – sees practically *nothing* as coincidence. If the cab-driver says "nice day isn't it?", and the schizophrenic is having a bad day, the latter will see the cab driver as mocking him and, perhaps, as being part of some conspiracy to make his day worse.) Instinctively people are very liberal at drawing connections and looking for common causes. Given this, and given how incredibly obvious the anatomical similarities between us and (e.g.) dogs and monkeys are, it is hard *not* to believe that, at some level, people practically *leapt* to the conclusion that we share a common ancestry with them.

Also, humankind's conception of itself – or, rather, what it *puts forth* as its conception of itself - is discrepant with some painfully obvious realities. And that self-concept, it seems reasonable to hold, embodies a defensive attempt to ward off these painful and humiliating truths. First of all, we cherish the idea that we are morally superior to animals, and that we are driven not by base urges, but by ideals: by a rational and humane conception of what is right.

It is true, of course, that people are, to some extent, motivated by a desire to do good. But it is also obvious that the desire to do good, and to live the "right" way, plays a decidedly less important role than animal drives in what

people do. If you base your predictions of somebody's behavior on the supposition that he will always act out self-interest, then – leaving aside computational error on your part – the deviation between your predictions and the person's behavior will be minimal. On the other hand, if you base your predictions of a person's conduct on the supposition that he wishes to live the "right" way, the discrepancy between the predications and the facts will probably be total.

Of course, it would be simple-minded to suppose that people are motivated entirely by self-interest. Most people *do* have values and *do* wish to live lives that have moral worth. Very few people are pure pragmatists. But when we consider the psychology of human motivation – the ratio of self-interest to morality in our motives -- it is hard to see any real discontinuity between ourselves and animals.

Another painful blow to our self-concept has to do with the biological basis of love. Everybody wants to be in love with a special someone. And the phenomenon of being in love is seen *not* just as something pleasurable, but as the highest actualization of the human spirit. I know that *my* life is richer and better when I am with someone I love. Obviously I'm not alone in this. And it is no mystery why we see love as being so valuable – why, indeed, we see love as a *sine qua non* of a meaningful existence and also as something of the highest excellence.

But when we look at the anatomical basis of love, this sine qua non of the good life - -this highest actualization of the human spirit – takes on quite a different countenance. Biologically speaking, the sexual act is the purest expression of our erotic instincts. And, emotionally speaking, it is also the most intense way to express love. So it is surely an irony of major proportions that male sexual organ is also an excretory organ, and that the act of ejaculating – of consummating one's love for another – has an obvious

resemblance to the act of urinating. Also, love between two men involves *both* of our excretory organs. And, finally, the vagina, though obviously distinct from the anus, is anatomically *and psychologically* very close to it. There can be no doubt that many men see the vagina as a kind of para-anus: that is surely connected with the fact that many love to perform anal sex with women. And, as a matter of empirical fact, many women think of their vaginas as dirty and shameful because, at least in part, of the proximity to, and resemblance of, the anus.

No function besides excreting makes as liberal use of the excretory organs as the sexual act. And yet this act is the most fulfilling thing a person can do and is also, supposedly, the consummation of all that is best about humankind.

Note that that two people on a date will feel embarrassment when they see dogs fucking – embarrassment because they know that what *they will* be doing – in the highest moments of love – is no better. Note also that dogs sniff women's genitals, indicating that *dogs* see us as fair game and thus as of their own ilk. Finally, note how ambiguous are our attitudes towards. Though sex, at least in certain contexts, is believed, supposedly, to be the highest manifestation of the human condition, it is also, evidently believed to be the lowest. It is only when performed under narrowly defined conditions, and thus when hyper-contextualized, that it is regarded as something holy, as opposed to profane. The need for those contextualizing is the proof positive that we are animals and that Darwin is right. Given that, for as long as civilization has existed, sexual intercourse has always been regarded as utterly indecent, except when done under such specific circumstances, shows that, as long as civilization has existed we have been aware of the fact that we are animals. And given the additional fact that a tribe's totem is almost always an animal, this awareness has always been coupled with an awareness

that we are descended from animals. This is confirmed that virile men are described as “tigers” and “stallions” and that attractive women are described as “vixens” (female foxes) and the like.

The Analogue/Digital Distinction and the Cogency of Kant's Transcendental Arguments

I. By a "proposition" I mean the non-linguistic entity meant by a complete sentence. I do not mean the sentence itself. "La neige est blanche" and "snow is white" mean the same thing. This common meaning is what I will refer to as a "proposition".

Kant held that some propositions are both *synthetic* and also *a posteriori.* (We will soon define these terms.)This contention is the heart of the system of thought set forth in the *Critique of Pure Reason.*

The purpose of this paper is to examine this contention of Kant's. I will argue that it is false and that Kant's argument for it is spurious. But I will also argue that Kant's false contention can readily be turned into a correct one and that his spurious argument can readily be turned into a cogent one. Nothing more than a few terminological changes are needed to accomplish this. Let me now outline the arguments I will give.

First, I will argue that the premises of Kant's argument turn out to entail the *opposite* of the thesis that there are synthetic a priori truths. Those premises entail that there are virtually *no* a priori truths, let alone *synthetic* a priori truths. Kant turns out to be even more of an opponent of a priori truth than John Stuart Mill.

Second, I will argue that Kant's reasoning involves a confusion concerning the two different epistemological roles that mental imagery can play. On the one hand, mental imagery can serve as a vehicle for conceptual analysis. On the other hand, mental images are spatiotemporal phenomena in their own right, and can thus be the *object* of empirical study. Kant fails to

distinguish the two quite distinct roles that mental images can have; and his views concerning the synthetic a priori are a reflection of that.

But I will also argue that, despite the problems just discussed, the really important parts of the *Critique* are a success. Kant's astonishing claims about causality, personal identity, and rationality turn out to be correct and founded on cogent arguments. But if this is to be made apparent, Kant's reasoning must be "de-psychologized": his statements about psychology must be reinterpreted as statements about logic. For the most part, this can be done by making a few adjustments to Kant's choice of words.

§ Let us begin with a point about terminology. The proposition *triangles have three sides* is necessarily true; there is no possible circumstance where it is false. Its truth is guaranteed by the structures of the concepts composing it – the concepts *triangle*, *three*, and so forth. It is clear that, even if there were no sentient beings, triangles would still have three sides, and the proposition *triangles have three sides* would still be true. [1] So in this context, the word "concept" obviously does not refer to a mental entity of any kind, even though that is what it refers to elsewhere. So in this context, the word "concept" refers to *non*-psychological, presumably platonic, entities of some kind. When I describe a proposition as "conceptually necessary", I mean that it holds entirely in virtue of facts about concepts *in the non-psychological sense of the word.* I thus will use the term "conceptually necessary" to denote the non-psychological necessity characteristic of *triangles have three sides.* Sometimes I will use the expression "conceptually true" as a synonym for "conceptually necessary.

As we will see, Kant recognizes a different category of necessary truth. In his view, some propositions are necessarily true because of facts about our minds. When I wish to discuss *that* kind of necessity, I will use the term

"representational necessity."

II. The first thing we must do is to say just what is meant by the terms "synthetic" and "a priori" and, therefore, what is meant by the conjugate terms "analytic" and "a posteriori."

Consider the proposition *1+1=2*. To know that is it true, one doesn't do empirical work; one doesn't perform experiments or make observations. *1+1=2* is thus a priori.

If a proposition is known to be true in some way *other* than through empirical work, Kant describes it as "a priori."

For Kant, a priority is an *epistemological* category. A priority is a property of our *knowledge* of propositions. Of course, Kant does describe propositions themselves as "a priori." But when Kant does so, he is really saying something about our *knowledge* of those propositions.

For Kant, a proposition is "a posteriori" if it is not a priori. The proposition *Smith weighs over 200 lbs* is *not* a priori; it is a posteriori. To know that is true, one must do empirical work.

Now let us turn to the expressions "analytic" and "synthetic." Here we must be careful. These days the term "analytic" is often meant to describe sentences. A sentence is "analytic" exactly if its semantics guarantees its truth. So "triangles have three sides" is analytic because, given what it means, it must be true. But this is not how Kant was using that term.

Kant uses the term "analytic" to describe propositions – not sentences, and not our knowledge of anything. Kant's official definition of "analytic" is this: a proposition is analytic if "nothing is contained in the predicate which is not contained in the subject." So *all tall males are males* is analytic because nothing is "contained" in the predicate ("males") which is not "contained" in the subject ("all tall males"). A proposition is "synthetic" if it is not analytic.

As Quine pointed out, there are two problems with Kant's definition of analyticity. First, the notion of containment" is "left at a metaphorical level."[2] Physical objects can contain other physical objects; but meanings don't contain meanings, at least not in the same sense. Second, Kant's formulation only applies to statements having subject-predicate form.[3] So it isn't immediately obvious how to apply it to statements not having that form, for example: *A=A* or *if p implies q, then p implies anything that q implies.*

But it isn't hard to adapt Kant's formulation to meet these two objections. Consider the proposition: *a+b is larger than a.* This proposition is composed of various concepts – *a+b, larger than,* and so on. By themselves, the structures of these concepts guarantee that *a+b is larger than a* is true.

a+b is larger than a is in a very different category from *John and Mary like to play tennis.* Consider the concepts that compose the latter proposition – *John, Mary,* and so on. The structures of these concepts do *not* guarantee the truth of that proposition.

For Kant, *a+b is larger than a* is a paradigm case of an "analytic" proposition, and *John and Mary like to play tennis* is a paradigm case of a proposition that Kant would describe as "synthetic". These facts suggest that, for Kant, a proposition is analytic exactly if it is "conceptually true" – exactly if it holds entirely in virtue of facts about the structures of concepts. It must be kept mind that, in this context, the concepts in question are not psychological entities.

It is easy to find support for this reconstruction of Kant's view. Consider the proposition *A=A.* [4] Kant describes this very proposition as analytic. But in so doing, he is not being true to his official definition of analyticity, namely "the predicate is contained in the subject". There is no clear sense in which this proposition has subject-predicate form, and there is no clear sense

in which any constituent of this proposition contains any other constituent. So if Kant's conception of analyticity were expressed by the definition just mentioned, then it would be inexplicable why he described *A=A* as analytic.

But in light of what we've said, it is clear what is motivating Kant. How is *A=A* different from *John and Mary like to play* tennis? The first is conceptually true; the second is not. In each case that Kant describes a proposition as "analytic", he believes that proposition to be conceptually true. In each case that Kant describes a proposition as "synthetic", he believes it *not* to be conceptually true. We may conclude that, for Kant, a proposition is "analytic" exactly if it is conceptually true – exactly if it holds entirely in virtue of the structures of the concepts composing it. A proposition is synthetic exactly if it is not analytic.

Now we can say what Kant means when he describes a proposition as both "synthetic" *and* "a priori." A proposition is a priori if it is known in some way other than through empirical work. A proposition is synthetic if its truth is not guaranteed by the structures of the concepts composing it. A proposition is synthetic a priori if it has both these properties.

III. According to Kant, examples of synthetic a priori truths are: *7+5=12* and *a straight line is the shortest distance between two points. 7+5=12* is a priori, in Kant's view, because no sense-experience is needed to know that it is true. It is synthetic because the concepts composing it do not, by themselves, make it true.[5]

Why does Kant think these truths are synthetic? Surely the truth of *7+5=12* is guaranteed by, and follows quite transparently from, the constitutions of the concepts 7, 5, +, and so on. So off-hand it is hard to see how that proposition could qualify as synthetic. Of course, 7+5=12 is not nearly as *obvious* as, say, *1+1=2*. But, apart from that, the two propositions

seem to be comparable.

Here we must consider an argument made by Arthur Pap. Kant held that *1+1=2* is analytic. In Kant's view, one cannot think *1+1* without thinking *2*. So given Kant's conception of analyticity, *1+1=2* must be classified as analytic – and this is exactly how Kant classifies it. *7+5=12* can be shown to follow analytically from *1+1=2* along with some other premises that Kant would regard as analytic.[6] Surely anything that follows analytically from an analytic proposition is itself analytic. So given his premises, Kant *should* regard *7+5=12* as analytic, even though he doesn't. Thus Kant's is guilty not only of error, but of inconsistency with his own premises, in believing that *7+5=12* is synthetic.

What about *a straight line is the shortest distance between two points*? It is hard to see how *this* could be synthetic. To avoid confusing mathematical technicalities, let us qualify the proposition in question. Let us consider the proposition *in a Euclidean space a straight line is the shortest distance between two points*. I don't know if this proposition is *obvious*. But given that it is true, it is hard to see how anything other than the constitution of the concepts involved could make that proposition be true.

What is Kant's reason for regarding these propositions as synthetic? To begin with, when you think *7+5* you don't necessarily think *=12*. Kant is right about this. When you think *m+n*, you don't necessarily think the sum of that operation. (That is why we need calculators.)

But at this point, Kant's reasoning becomes fuzzy. The fuzziness is due to the fact that Kant was not clear about whether he is using the term "concept" in the logical, or the psychological sense; he seems to go back and forth. It does not appear that he registered this distinction. But I think it is fair to reconstruct his reasoning as follows.

Given that you don't think *=12* when you think *7+5*, it seems to follow

that *your* concept of *7+5* doesn't contain your concept *=12*. So in exercising your grasp of the concept *7+5*, you are not exercising your grasp of the concept *=12*. Therefore your grasp of the concept *=12* is altogether distinct from your grasp of the concept *7+5*.[7]

When you think *straight line*, you don't necessarily think *shortest distance between two points*. What you *think*, says Kant, is purely *qualitative*, not quantitative; you think of a line with a certain shape, not a line of a certain distance. So your grasp of the concept *straight-line* is distinct from your grasp of the concept *distance*.[8]

From this Kant apparently infers that *the* concept *7+5* is entirely distinct from *the* concept *=12*. But this inference involves a muddle of the two meanings of the word "concept". *My* concept of *7+5* is distinct from *my* concept of *=12*. But this doesn't show that *the* concept *7+5* is distinct from *the* concept *=12*.

For Kant to make his case, he needs to appeal to some principle like the following: the structure of concepts (in the objective sense) of the word can be read off of the way in which they are represented in our minds. Facts about our *grasp* of concepts must parallel facts about the concepts themselves. So if my *grasp* of *7+5* doesn't comprise my *grasp* of *=12*, then the first concept doesn't comprise the latter.

Once that principle is granted, Kant's argument goes through. But why grant it? What is the motivation for it? That principle presupposes that concepts are completely transparent: you cannot grasp a concept without at once grasping everything definitive of it. There is no more to them than meets the eye. Their *esse* lies in their *percipi*. It thus seems to me that the principle in question presupposes, and is probably motivated by, a kind of anti-realism about concepts – by a kind of psychologism concerning them.

They are creatures of the mind, not entities in their own right.

Let us now close Kant's argument. As we've seen, in Kant's view, the truth of *7+5=12* does not follow *wholly* from the concepts involved in that proposition: it is therefore synthetic. At the same time, Kant denies that it is an empirical truth: he denies that it is made true by what happens when apples are laid next to apples, when quarts of liquid are added to other quarts of liquid, and so on. Arithmetical truths are not inductive generalizations that we could, in principle, find to be false. You don't know them through empirical work: you know them *a priori.* So such truths are a priori *and* synthetic. For analogous reasons, the same is true of *a straight line is the shortest distance between two points* and any other proposition that is non-empirical but (unlike *A=A*) non-trivial.

The very existence of synthetic a priori propositions raises difficult questions. Such propositions are not made true by facts about concepts. At the same time, they are not made true by the kinds of facts that can be discovered through empirical investigation. There doesn't seem to be anything left to make them true. So what, in Kant's view, makes a synthetic a priori be true? And how do we know such propositions be true?

IV. Let us now give Kant's answers to the questions just posed. How do you know that 7+5=12? Not through conceptual analysis. Not through the senses. You learn it through an act of the imagination. You must *imagine* seven objects laid next to five other objects, and then see that the resulting aggregate comprises exactly 12 objects.

How do you know that a straight line is the shortest between two points? Not through conceptual analysis. Not through the senses. You must have a certain kind of para-sensory experience: you must *imagine* a line connecting two dots.

Kant refers to images of this sort as “intuitions”. So it is through “intuition” that you learn the truth of *7+5=12*.

It must be made clear that Kant is *not* using this term in the way it is used today. These days, philosophers usually use the term to refer to a putative insight into concepts (“my intuition is that killing someone is morally worse than passively allowing him to die”). This is not what Kant means. By “intuition” Kant means an act of *imaging*: a kind of para-perception.[9]

Let us continue with Kant’s story. How exactly do these para-perceptual experiences reveal the truth of *7+5=12*? The idea seems to be as follows. First of all, these para-sensory experiences don’t show *conceptual* or *logical* necessity. Kant’s whole point is that the *concepts* involved in 7+5=12 do not, by themselves, guarantee its truth. Taken by themselves, those concepts are consistent with the truth of 7+5≠12. So the para-perceptions that teach us the truth of 7+5=12 do not do so by apprising us of necessary relations holding among the components of that proposition.

At the same time, those para-sensory experiences don’t tell us anything about the external world: you can’t learn anything new about the external world through an act of imagination.

So, for Kant, these para-perceptual experiences show neither empirical truth nor conceptual necessity. What they show is *representational necessity*. They show us facts, not about the world or about concepts, but about how we must *sense-perceive* the world.

These para-perceptual experiences show us what we can and cannot *imagistically* represent. We cannot represent *in an image* a straight line between two points that is longer than some other line between those two points. We cannot form *an image* of four apples that isn’t also an image of two apples and two other apples.

Let us sum up. Recourse to “intuition” apprises us of what we can, and

cannot, *imagistically* represent. So these "intuitions" apprise us of what we can, and cannot, represent in *perception*; they tell us how we must *see* things, in the literal sense of "see."

At this point, we must be careful. One might be tempted to ground these "representational" facts in conceptual facts:

There can't be an image of four apples that isn't an image of two apples and two other apples because the *concepts* 2, + etc. are such that they don't permit this. It is a *conceptual* or *logical* truth that 2+2=4, and this fact underlies the representational limitation Kant has in mind.

This may be true. (In my view, it *is* true.) But it is just what Kant cannot, and does not, say. He thinks that *7+5=12* is such that its truth is *not* guaranteed by the concepts composing that statement. So the representational limitations he has in mind do *not* have a purely conceptual or logical foundation.

In Kant's view, these necessities are really of a *psychological* nature. Our "forms of thought" are such that we must see the world (literally *see* it) as being such that, in it, parallel lines don't meet, a straight line is the shortest distance between two points, two objects coupled with two other objects make four objects, and so on. We are wired in such a way that *we* (though perhaps not other life forms) must see objects as being in a space and time with a certain structure: a structure that does not permit parallel lines to meet, that does not permit 2+2 to equal 3, and so on. So "intuition", i.e. para-perceptual experience, shows us how we must sense-perceive the world.

IV. Is there any room in this picture for synthetic a priori *truth*? And supposing that it does have room for it, does it have room for *knowledge* of

it? Both questions must be answered with a “no”. My argument for this owes much to Laurence Bonjour.[10]

For the sake of argument, let us concede a couple of key assumptions to Kant. First, let’s concede that we *do* sense-perceive the world as being such that 7+5=12, a straight line is the shortest distance between two points, parallels don’t meet[11], and so on. Second, let’s concede that we learn these facts about how we perceive the world through “intuition” (some kind of para-perceptual experience), and *not* through conceptual analysis or through sense-perception.

Given only that we are psychologically compelled to *see* the world as being such that, in it, thus and such is the case, it doesn’t follow that thus and such really is the case. A psychotic may be psychologically compelled to see pink elephants, but that doesn’t mean there are pink elephants. So our para-sensory experiences don’t tell us anything about the world, but only (at most) about how we perceive it. They tell us *not* that 7+5 *really* equals 12, but only that *we* are psychologically compelled to sense-perceive the world as being such that, in it, 7+5=12.[12]

Actually, Kant himself admits this. These intuitions, he says, tell us about the “phenomenal” world, not the “noumenal” world. By Kant’s own definition, the “phenomenal world” is just the world as it appears to us: it is how *we see* the world. The “noumenal” world is the actual world. So by Kant’s own admission, these intuitions inform us only about how we see the world, not about how it is. We must see the world as being such that, in it, parallels don’t meet. But whether this is true in the “noumenal” world is a totally separate issue. (Kant says that we simply cannot know whether they meet in the noumenal world or not.[13]) So intuition doesn’t really tell us

anything about the external world.

At the same time, for Kant, intuition doesn't tell us anything about the structure of concepts either. Intuitions are *not* insights into concepts. Kant is adamant about this. Intuition begins precisely where conceptual analysis ends.

So, on Kant's view, intuitions don't tell us anything about the spatiotemporal world and they don't tell us anything about concepts. Thus, given only that intuition tells us that a straight line is the shortest distance between two points, we simply have no reason to believe that straight line *is* the shortest distance between two points. At most, we can conclude that we are psychologically compelled to *see* the world as being such that, in it, a straight line is the shortest distance between two points.

So "intuition", on Kant's view, doesn't give us *knowledge* of the fact that 7+5=12. Intuition gives us only the knowledge that we are psychologically compelled to see the world as being such that, in it, 7+5=12 and parallels don't meet. *A fortiori* intuition doesn't give us *a priori* knowledge of the fact that 7+5=12. If it gives us a priori knowledge of anything, it is of the fact that we must *perceive* the world as being such that, in it, 7+5=12. But, as we'll see in a moment, it doesn't even give us a priori knowledge of this; it gives us, at most, *a posteriori* knowledge. [14]

V. 7+5=12 obviously *is* true. So something makes it true. For Kant, facts about concepts don't make it true. As we've seen, facts about our own psychology don't make it true. The only thing left to make it true is the spatiotemporal world. So on Kant's view, followed to its logical conclusion, for us to know whether 7+5 *really* equaled 12, we'd have to learn about the actual world (which, for reasons that don't concern us, Kant says we cannot do). So it becomes a kind of empirical truth. In any case, it becomes a truth

that *cannot* be known in any way other than through examination of the external world. So, by any definition of *a priori*, it is not a priori.

The same thing holds of *any* truth that we would characterize as "a priori", apart from completely trivial ones (e.g. *1=1*). If they can be known at all, it is through knowledge of the noumenal world. According to Kant, we can't know anything about the noumenal world. Therefore we don't have *any* knowledge of such truths, let alone a priori knowledge.

VI. One might make the following claim on Kant's behalf:

How could we learn the truths of geometry, if not through visual intuitions? In any case, whether or not we *could* learn them in some other way, we *do* learn at least some of them through intuition. This would seem to support Kant's view.

When you learn through mental imagery that a straight line intersects a circle at no more than two points, you are learning something that is made true by the *concepts* CIRCLE and LINE, and so forth. You are not learning something that is made true by our cognitive machinery. Nothing about circles could be made true by our cognitive machinery. When we *do* learn facts about geometry or perhaps arithmetic through an exercise of powers of imaging, what is really going is *conceptual analysis*, the vehicle for which, in those cases, happens to be mental imagery. There is no reason why images should not mediate conceptual analysis. So given only that some truth is discovered through thought involving mental imagery, it doesn't follow that it is synthetic.

What led Kant astray was a failure to distinguish the different roles that mental imagery can play in the acquisition of knowledge. Kant noted that

mental imagery is a kind of para-perception. He concluded from this that, when mental imagery is involved in the discovery of some truth, that truth is synthetic – like those truths that we use *perception* to discover. The idea is that, if the truth in question were purely conceptual, then we wouldn't need to use para-perception to ascertain it.

But that is simply false. If somebody asks me "what is the minimum number of sides a solid can have?", I base my answer on a mental image that I produce.[15] I *imagine* trying to produce a solid out of four sides, then note that it doesn't work, then try to do the same thing with five sides, and note that it *does* work. But surely the fact that a solid has to have a minimum of five sides is not synthetic. That proposition holds entirely in virtue of mathematics – in virtue of the structure of the concepts *five* and *solid* and so forth. It is true that I rely to some extent on imagery to learn this truth. But the imagery is a vehicle for *conceptual* analysis.

It might seem that Kant *did* show that we have *some* a priori knowledge: we have *a priori* knowledge of facts about the phenomenal world, i.e. about how we must perceive the world. We know a priori that the world, as we perceive it, is such that parallel lines don't meet in it, triangles have interior angles adding up to 180°, and so on.

The truth is that, on Kant's account, even *this* knowledge turns to be a posteriori. Once again, what led Kant astray was a failure to distinguish the two very different epistemic roles that attention to our own mental processes can have.

As we just noted, mental imagery can sometimes serve as a vehicle for conceptual analysis. But it can also play a very different epistemic role. Suppose I want to find out facts about my own psychology; suppose, for example, I want to find out if I can imagine a 17-sided figure. Here I will, of course, generate mental imagery, and attend to the content of that imagery.

But in this case, the imagery is not serving as a vehicle for conceptual analysis. In attending to my mental imagery, I am actually doing *empirical* work, and the truth that I discover (*I cannot visualize a 17-sided figure*) is as a posteriori as *Socrates was bald.* Here the mental imagery is the actual *subject-matter* of the truth-learned.

For Kant, you learn that you can't represent parallel lines as intersecting by *noting that you cannot imagine it;* you *observe* your own mental imagery. So you learn this putative fact about your representational powers through a kind of self-observation, and not through any kind of conceptual analysis. This knowledge turns out to be quite a posteriori. When Kant says that "in the phenomenal world" parallels don't meet, he is just saying that we cannot visually represent a world where parallels don't meet. So our knowledge of this fact about the phenomenal world turns out to be a posteriori.

Consider a point made by Leibniz about the role that imagery plays in the acquisition of mathematical knowledge:

> But I do not agree with you what seems to be your [Locke's] view, that some kind of general certainty is provided in mathematics by "particular demonstrations" concerning the diagram that has been drawn. You must understand that geometers do not derive their proofs from diagrams, although the expository approach makes it seem so. The cogency of the demonstration is independent of the diagram, whose only role is to make it easier to understand what is meant and to fix one's attention. It is universal propositions, i.e. definitions and axioms and theorems which have already been demonstrated, that make up the reasoning, and they would sustain it even if there were no diagram.[16]

Diagrams serve as *vehicles* for a priori reasoning. The fact that I use a

diagram to prove the intermediate value theorem or the Pythagorean theorem doesn't prove that either of those theorems is synthetic. They aren't synthetic. This point applies to mental imagery – to mental diagrams.

Leibniz's point was made more recently by Christopher Peacocke:

When you come to know a logical truth by way of your having a proof of it, you may need to perceive the inscription of the proof, and you may need various perceptual capacities to appreciate that it is a proof. But the justification for your belief in the logical truth is the proof itself. Perceptual experience gives access to the proof, which provides an experience-independent justification for accepting its conclusion. By contrast, if you come to believe "That's Mikhail Gorbachev", when you see him at the airport, what entitles you to your belief is (in part) the perceptual experience by which you recognize Gorbachev. Your perceptual experience is not a mere means which gives you access to some experience-independent entitlement to believe "That's Gorbachev." [17]

When you use a mental image to figure out whether a line can intersect a circle at more than two points, the image *gives* you the justification for believing a certain proposition, but the image is not *itself* a justification-- at least not in the sense in which seeing Gorbachev at the airport is justification for believing that Gorbachev is still alive.

VII. I should respond to a possible misgiving about our analysis:

You have belittled and caricatured Kant's views. Kant says we must see the world as being such that, in it, parallel don't meet, 7+5=12, and so on; and he says that this due to our forms of thought. You have taken this to mean that, for Kant, our seeing the world as having these features is due to some merely psychological compulsion, like a neurotic's compulsion to wash his hands twenty times a day. But this is to misunderstand the nature of the necessity Kant has in mind. The necessity that attaches to the proposition *parallels don't meet* is epistemological, not psychological; transcendental, not empirical. That proposition describes a necessary *precondition* for our being able to have any experience at all of the world. Obviously *parallels don't meet* won't be made true by a mere psychological compulsion, just as pink elephants aren't brought into existence by a psychotic's hallucinations. But Kant in no way holds that the necessity attaching to *parallels don't meet* is psychological. In fact, he goes out of his way to make it clear that the necessity in question is non-psychological. In characterizing what is in fact transcendental necessity as mere psychological compulsion, you have misrepresented Kant's view, and made it far easier to destroy than it really is.

In the end, "transcendental" necessity reduces *either* to mere psychological necessity *or* to conceptual necessity. We cannot coherently think of a world where triangles have two sides or where 2+2=3. That is because propositions like *2+2=3* and *triangles have two sides* are *conceptually* false and consequently draw a limit to what *we* can conceive of. What cannot be *conceived* cannot be *per*ceived.[18] Conceptual necessities thus impose sharp limits on how the world can be visually represented. So it is clear how conceptual falsity translates into representational impossibility.

At the same time, if a proposition is conceptually possible, then it is hard

to see how it could ground a representational impossibility. Because we have certain cognitive machinery -- certain "forms of thought" -- we are compelled to see the world in a certain way. But unless those forms of thought embody *conceptual* necessities of some kind, I don't see why we couldn't, in principle, have different forms of thought.

An example might help. We could obviously imagine having a cognitive architecture such that we could only see two different colors or hear two different pitches. Under those circumstances, given the forms of thought operative in our sense-perception, we simply *couldn't* see a third color or hear a third pitch. But obviously it is *conceptually* possible that there be more than two colors. (It is actual; so it is possible.) So it is representationally possible.

There are limits to what our perceptual machinery can do. But it is very hard to see how these limits could be *necessary* unless they were grounded in *conceptual* impossibilities.

VIII. There is an incoherence in Kant's thinking even more radical than any mentioned so far. He rightly wants to hold that propositions like 7+2=9 and *the area of a (Euclidean) triangle is ½ the base×height* are a priori. But he also says that we *sense-perceive* the world as being such that these propositions are true *in it*. Of course, if the world can be *seen* (literally *seen*) as being such that 7+2=9 is true *in* it, then 7+9=12 is *descriptive* of the spatiotemporal world. If *7+2=9* said absolutely nothing about the spatiotemporal world, then the truth or falsity of that proposition couldn't make any difference to how the spatiotemporal world seemed to us. So, on Kant's view, these propositions say something about the spatiotemporal world.

But *7+2=9* doesn't say anything about the spatio-temporal world; it

doesn't say that if you hold up two fingers and then hold up seven fingers, nine fingers are on display. For there could, in principle, be a causal law that falsified just that sort of statement – a law that didn't allow more than five fingers to be simultaneously on display in a certain region.

The same is true of Kant's claim that we must sense-perceive the world as being such that, in it, straight lines are the shortest distances between pairs of points. The proposition *a straight line is the shortest distance between two points* says nothing at all about the spatio-temporal world; it defines a certain kind of space. Of course, there are empirical propositions – concerning light rays and point-masses – that are analogous to that purely mathematical proposition. But surely Kant doesn't have these in mind when he says it is *a priori* that a straight line is the shortest distance between two points.

So Kant's position is self-undermining. He needs mathematical propositions to be two things at once. They must be descriptive of the world: they must say something about beams of light. But they must also *not* be descriptive of the world: they must not be capable of falsification by anything having to do with beams of light and the like.

A Kantian might reply by saying:

Propositions like *7+2=9* and *a straight line is the shortest distance between two points* do describe the empirical world. They describe *necessary* features of it. So *7+2=9* is like *grass is green* in one respect: it is descriptive of the empirical world. But *7+2=9* is unlike *grass is green* in another respect: it describes *necessary* features of the world. That is why *7+2=9* is a priori, unlike *grass is green.* There is no inference from *P describes the world* to *P is an empirical and contingent proposition.* So the point you just made fails.

But there *is* an inference from *P is necessary*[19] to *P doesn't describe the empirical world (in any way at all).* If P is necessary, then P's being true in world 1 doesn't distinguish world 1 from world 2 or from world 897. P therefore doesn't say anything about the world that distinguishes it from other possible worlds. So P doesn't say anything at all about the world. A "description" of a thing that fails to distinguish that thing from other things in any way at all is no description at all.

IX. So far I've said a lot about why Kant is wrong. But I think that, with a few adjustments, Kant's position might possibly be correct.

Ultimately, there is one thing wrong with Kant's view: his conception of analyticity is much too narrow. Everything else that is wrong about Kant's position can be traced back to this one mistake.

Kant grants analytic status *only* to propositions like

(a) circles are round.

He *denies* analytic status to propositions like

(b) a circle is a class of points equidistant from a given point in a plane.

Thus he says the following is analytically false:

(c) circles are not round.

But he *denies* that the following is analytically false:

(d) a circle is *not* a class of points equidistant from a given point in a plane.

His reasoning seems to be this:

A proposition is analytic exactly if one is guilty of incoherence in virtue of countenancing its negation. One is guilty of incoherence if one thinks that (*) is false. But is not guilty of incoherence if one thinks that (**) is false. One is *wrong* to think that (**) is false, but not incoherent.

To be incoherent is to think in a way that is counter-conceptual. Thus (*) is counter-conceptual. (**) is not counter-conceptual.

An analytic proposition is one that is made true by concepts. An analytically false proposition is one that is made false by concepts: a counter-conceptual proposition. One can countenance the negation of (**) without being guilty of incoherence, and without countenancing a counter-conceptual proposition. Thus (**) is not analytic.

But (**) is a priori. So we must suppose that it is made true by facts about our conceptual machinery; it embodies a fact about our conceptual schema. That is why no empirical discovery could confute it.

The word “coherent” has two entirely different meanings: a psychological meaning and a purely logical meaning.

If you believe (d), you are guilty of *logical* incoherence: the proposition you believe entails a contradiction. But a perfectly *sane* person could believe the negation of (d). In virtue of believing it, one is perhaps guilty of stupidity, and of incoherence in the *logical* sense, but not of incoherence in the psychological sense.

By contrast, if you believe (c), you *are* guilty not just of logical, but of psychological, incoherence. You couldn’t possibly think that *circles* are not

round; to think of something as a circle just *is* to think of it as round. So to think of something as a non-round circle – if such a thought is even possible -- would involve a kind of double-think, a fragmentation of the psyche.

The following story might be helpful. Somebody tells you:

(e) a triangle is the area bounded by three straight lines such that any two of them intersect but not all three intersect.

But you don't immediately see the truth of that proposition. Initially, you reject you it; you say:

(f) A triangle is *not* the area bounded by three straight lines such that any two of them intersect but not all three intersect.

Are you guilty of *psychological* incoherence? It seems not. A perfectly sane person could believe (f). Belief in (f) constitutes mathematical, but not *psychological* incoherence.

But a sane person could *not* believe:

(g) A triangle is *not* a three sided figure.

If one believed (g), one would be guilty not only of logical, but of *psychological* incoherence. Thinking of x as a triangle *and* as not being three-sided would involve a fragmentation of the thinker's psyche.

Logical coherence is a relation among concepts. Psychological coherence is a relation among our *grasp* of concepts.

All of the following are analytic:

(a) circles are round.

(b) a circle is a class of points equidistant from a given point in a plane.

(e) a triangle is the area bounded by three straight lines such that any two of them intersect but not all three intersect.

(h) triangles have three sides.

But (a) and (h) are different from (b) and (e). One is guilty, not only of logical, but of *psychological* incoherence in virtue of believing (a) and (h). One is guilty of logical, but not of psychological incoherence, in virtue of believing (b) and (e).

Let us set say that (b) and (e) are "hard analytic", and let us say that (a) and (h) are "easy analytic". It is *hard* to know that (b) and (e) are true. It is not hard to know (a) and (h) are analytic; a grasp of those propositions by itself suffices for a knowledge of their truth. (Kant describes such propositions as "apodictic.")

What Kant refers to as "synthetic a priori" truths are nothing other than what we are describing as hard analytic truths. What Kant refers to as analytic truths are what we are describing as easy analytic truths.

Once it is recognized that (b) and (e) are analytic, albeit *hard* analytic, we must say that they hold entirely in virtue of facts about the concepts that compose them; and there is no longer any temptation to say that they are made true by facts about the mind. So that obliterates any possible warrant that might otherwise be had by Kant's psychologistic analysis of those propositions.

§ But we must remember what Kant's larger concern is. Kant's main concern is *not* a taxonomy of propositions. That is incidental to what he is doing.

We have experience; we have sequences of thoughts and perceptions that cohere. The world is intelligible to us. We sense-perceive it; we think about it; and, up to a point, our perceptions and thoughts are accurate. In Kant's view, given the very fact that these things are true, we can deduce certain things about the structure of the world. Given *only* that we have coherent ideation, the world must satisfy *certain* minimal conditions.

Kant refers to those conditions as "transcendental": they are the conditions that are necessarily satisfied by any world in which there are entities that have coherent (if not always accurate) thoughts and representations. The "necessity" just mentioned is not causal: it is of a deeper kind.

A "transcendental argument" is one having the form: given only that we have experiences of such and such a kind, it follows that the world must have thus and such features.

Kant gives transcendental arguments for many things. I think that some of these arguments are cogent, *provided* that we make some terminological changes. Where Kant talks about "synthetic a priori", we should instead talk of "hard analytic". This minor adjustment de-psychologizes Kant's arguments, and makes them go through.

Consider the fact that we have coherent sequences of mental representations. That fact *does* show that the world satisfies certain very basic conditions. I will argue that there is an *analytic* connection – a *hard* analytic connection, but an analytic connection no less – between the proposition *there exist beings who have the cognitive wherewithal to raise metaphysical questions* and propositions concerning the structure of the

world itself.

I will argue for another significant claim. Consider the following propositions: *any two events are related in space or in time; there are causal relations among states of affairs; objects are not mere bundles of properties, but are things in which properties inhere; human beings are not mere bundles of perceptions (or thoughts), but are things in which such entities inhere.* These propositions describe the world. I will argue that, just as Kant maintained, each of those propositions is entailed by the proposition *there exist beings who have the cognitive wherewithal to raise metaphysical questions.*

§ One of Kant's main objectives is to show that there are necessary connections among the events that we experience. Hume said that there are sequences and regularities of events, but there is never any kind of internal connection holding among them. For event x to cause event y is just for that sequence of events to instantiate a general regularity.

Hume's analysis is counter-intuitive. Intuitively, it seems that pushing the button *makes* the computer turn on, that flipping the switch *makes* the light turn on. Understandably, Kant wanted to do justice to that intuition. He tries to justify this intuition with a transcendental argument. (Here I will give *a* transcendental argument for that view; whether it is Kant's own argument, or even one he would consider credible, is an exegetical question I wish to avoid.) Given the very fact that we are even here to raise the question whether there are "inner connections" among events, it follows that *there are* such connections. Imagine a world where there is *no* causation, where there are merely sequences of events. Such a world would comprise nothing enduring. Persistence is itself a causal notion. If something persists, there is *ipso facto* a causal sequence of some kind. So where there is no causation,

there is no persistence. Where there is no persistence, there is no coherent ideation. For surely, as a matter of necessity, a coherent thought process involves some kind of persistence. So given only the fact that there are beings who have the ability to raise questions about anything, let alone about the necessary preconditions for experience, it follows that causal relations do not consist in mere sequences of events, but in internal relations of some kind. [20]

§ Here is a similar argument. The very existence rational thought shows that, in addition to thoughts, there must things which *have* those thoughts. This contradicts Hume's contention that the self is a mere "bundle" of thoughts.

Suppose the following. I think *if P then Q* . Fred thinks *P*. Sam thinks *Q*. That sequence of thoughts is not a coherent or rational thought process. . Suppose I think *if P then Q*, Fred thinks *Q*, and Sam thinks *P*. That sequence of thoughts was not an *irrational* thought-process. Suppose *one* person thinks: *if P then Q, and P, therefore Q.* Here we have a case of rationality. Suppose *one* person thinks: *if P then Q, Q, therefore P*. Here we have a case of irrationality.

Hume says that there are *only* sequences of perceptions and thoughts; there are not things that *have* those thoughts. So a Humean cannot distinguish among the cases above. For there to be *rational* or *irrational* thought processes, there must be things which *have* those thoughts: there must be *people* beneath the thoughts. Only then will a precondition of rationality (or irrationality) be met. There obviously are such things as coherent and incoherent sequences of thoughts. Therefore, there are such things as persons. The very existence of coherent (and incoherent) thought processes shows something about the structure of the world: it shows that, if there are to be such things as thoughts, there must be things underneath the thoughts which

have them. So here we have a transcendental argument for the existence of selves.[21] In the same stroke, we have a transcendental argument for the existence of internal relations among distinct mental events.

§ But notice that these transcendental arguments turn on the existence of entirely *analytic* connections among propositions; no "synthetic a priori" is involved. Given the very *concept* of a rational thought process, it follows that not just *any* sequence of thoughts constitutes a case of rationality or irrationality: only *certain* sequences do so. Given the existence of coherent and incoherent thought processes, it follows *analytically* that there are things underlying thoughts which have them: it follows analytically that the Humean conception of person identity is false. The analyticity in question is *hard* analyticity, but it is analyticity no less.

Bonjour, Laurence. 1998. *In Defense of Pure Reason.* Cambridge: Cambridge University Press.

Burge, Tyler. 2000. "Frege on Apriority". *In New Essays on the A Priori* edited by Paul Boghossian and Christopher Peacocke Clarendon Press Oxford: 2000

Kant, Immanuel. 1965. *Critique of Pure Reason.* Translated by Norman Kemp Smith. St. Martin's Press New York.

Kuczynski, John-Michael. "A Solution to the Paradox of Causation."

Philosophy in Science. Volume 8. 1999.

Montague, Richard. *Formal Philosophy.* Edited by Richard Thomason. New Haven: Yale University Press.

Peacocke, Christopher. 2000. “Explaining the A Priori: The Programme of Moderate Rationalism”. *In New Essays on the A Priori* edited by Paul Boghossian and Christopher Peacocke Clarendon Press Oxford: 2000

Quine. 1990. “Two Dogmas of Empiricism.” In *The Philosophy of Language.* A.P. Martinich (ed.). Oxford: Oxford University Press.

Pap, Arthur. 1958. *Semantics and Necessary Truth.* New Haven: Yale University Press.

Reichenbach, Hans. 1958. *The Philosophy of Space and Time.* Translated by Maria Reichenbach. New York: Dover.

Can One Grasp Propositions Without Knowing a Language? Defending Cognitive Realism against Robert Brandom

RESUMEN

¿Debe uno conocer un lenguaje para captar proposiciones? De acuerdo con la opinión de muchos la respuesta es 'sí'. Generalmente la idea que está detrás de esta posición es que las proposiciones, siendo entidades platónicas, no pueden ser contempladas *directamente*: así uno debe captarlas a través de representaciones de las mismas. Y, por supuesto, se supone que las representaciones en cuestión son símbolos lingüísticos. Wittgenstein añadió una nueva dimensión a esta línea de pensamiento: defendió que el *lenguaje* y por tanto los símbolos lingüísticos, son esencialmente públicos: un lenguaje es un conjunto de símbolos gobernados por *reglas* y éstas son esencialmente públicas: esto se debe a que una regla 'privada' puede desobedecerse con impunidad y por esto vale tanto como si no fuese una regla. Así, en opinión de Wittgenstein, uno debe ser miembro de una *comunidad* para conocer un lenguaje y de este modo poder captar una proposición. Robert Brandom da un paso más allá y afirma que las *comunidades* son los únicos seres *realmente* capaces de captar proposiciones y que los individuos las captan sólo en un sentido derivado. En este artículo voy a defender que uno *puede* captar proposiciones sin captar un lenguaje y que la capacidad de aprender un leguaje presupone la capacidad para captar proposiciones. Se seguirá de lo que digo que son los individuos y no las comunidades los que de un modo genuino captan proposiciones.

ABSTRACT

Must one know a language to grasp propositions? According to many, the

answer is ‘yes’. Generally, the idea behind this position is that propositions, being platonic entities, cannot be beheld *directly*: so one must grasp them by way of grasping representations of them. And, of course, the representations in question are usually taken to be linguistic symbols. Wittgenstein added a new dimension to this line of thought: he argued that *language*, and thus linguistic symbols, are essentially public: a language is a set of *rule*-governed symbols, and rules are essentially public: this is because a ‘private’ rule can be disobeyed with impunity, and is therefore as good as no rule. So, in Wittgenstein’s view, one must be a member of a *community* to know a language and therefore to grasp a proposition. Robert Brandom goes a step further and says that *communities* are the only beings *truly* capable of grasping propositions, and that individuals grasp them only in a derivative sense. In this paper, I will argue that one *can* grasp propositions without grasping a language, and that the ability to learn a language presupposes the ability to grasp propositions. It will follow from what I say that it is individuals, and not communities, that genuinely grasp propositions.

I: INTRODUCTION

In *Making It Explicit* Robert Brandom contends that, in order to grasp *propositions* — that is, to grasp truths or falsehoods — a creature must be a member of some kind of community. In fact, Brandom holds a much stronger position: he says that intentionality can be ‘non-derivatively’ [Brandom (1998), p. 151] attributed only to communities of individuals, and only derivatively applied to individuals.[22] In taking this latter position, Brandom is of course opposing common sense: it is ordinarily thought that communities cannot think, or in any way be conscious, except in a metaphorical sense; and that to say of a community that it thinks a certain way or has a certain kind of ‘consciousness’ is really just to make some kind

of statement about the individuals belonging to it.

Why does Brandom hold that, in order to grasp propositions, an individual must belong to a community? Because, says Brandom, to be able to grasp propositions a creature must know a *language* [Brandom (1998), pp. 148-59] [23]; and, as Brandom also holds, a creature that isn't a member of some kind of community cannot possibly know a language [Brandom (1998), p. 155]. These views coincide with views held by the later Wittgenstein. Moreover, Brandom's reasons for holding these views are heavily derivative of reasons set forth by Wittgenstein. We will presently consider both Brandom's and Wittgenstein's reasons.

In this paper I wish to show that, in order to learn a language, a creature must *already* be able to grasp propositions and, more generally, that for a creature to *become* conscious of some norm it must *already* be conscious of norms. Consciousness of norms is either innate or non-existent. Of course, consciousness of this or that norm needn't be, and usually isn't, innate. But a creature that is not innately aware of some norms couldn't possibly *become* aware of any. Experience – so I will try to show -- apprises us of new norms only by articulating some innate connectedness to normativity in general.

I will also try to show that, given a group of creatures none of whom are aware of norms, there is no way that they could come to be aware of norms by interacting with one another; and there is no way that a group of creatures could collectively be aware of norms unless some of the individuals composing that group were previously aware of norms.

Finally, we will consider Brandom's arguments *against* the position just described. This will involve our considering Wittgenstein's arguments against that same position. We will find that both Brandom's and Wittgenstein's arguments involve identifiable fallacies.

This paper will defend a version of what is called 'cognitive realism'.

What does this term mean? Many of our abilities — e.g. our ability to understand language, to solve mathematical problems, to play musical instruments — *seem* to result from our having an 'implicit awareness' of certain rules. They seem to result from an awareness of such rules because our mental and behavioural productions can, without artifice, be made out to accord with such rules: there is a kind of isomorphism between these productions and such rules. And it would be an extraordinary coincidence if this isomorphism obtained without our cognitive productions' actually *resulting* from some kind of awareness of these rules. As a competent English speaker, I *will* say 'the dog chased the cat' and I certainly won't ever say 'dog cat chased the.' Why won't I ever say this? One hypothesis is: I have an 'implicit grasp' of certain grammatical rules; I have mastered rules governing what concatenations of sound betoken well-formed sentences of English; and according to these rules 'the dog chased the cat' is grammatical and 'dog cat chased the' is not grammatical. This kind of explanation would seem to apply *mutatis mutandis* to virtually every kind of cognitive competence. There is no shortage of thinkers who have taken this position, in at least some form — Chomsky, Dummett, Kant.

Chomsky and Dummett are 'cognitive realists': we really are aware of certain rules — these rules have 'cognitive reality' — and this awareness determines what we think and do. I will sometimes refer to this position as 'cognitivism'.

There is no shortage of thinkers who reject cognitivism. They usually argue as follows. First of all, given only that a process or event can be described *as though* it resulted from consciousness of some rule R, it doesn't follow that it *did* result from such consciousness. The movements that my fingers are currently making could be described *as though* they resulted from my following some kind of complex rule involving the relative distances of

various comets from one another. But this rule is not in fact operative. This is the first leg of the anti-cognitivist argument [Searle (1992) and (1999)].

The next leg is as follows. The idea of sub-personal cognitive activity is incoherent: so not only is it *not* de rigueur to see (e.g.) my visual perceptions or linguistic intuitions as resulting from such activity: it is de rigueur *not t*o see them that way. Why is the idea of sub-personal cognition held to be incoherent? At this point, different anti-cognitivists give different answers. But their various arguments tend to fall into one of a very few categories.

One line of attack is this: there can be no such thing as a 'private language.' Sub-personal cognition would make use of just such a language. (Cognition obviously involves the processing of information; and information must be mediated. Anything mediating information is *ipso facto* a language [Dennett (1978) and Fodor (1975)]. Obviously the sub-personal processes posited by e.g. Chomsky and Marr are not inter-personal: they are *intra*-personal. So they involve *intra*-personal, i.e. private, languages.) So such cognition is an impossibility.[24] [JK1] We will soon enough discuss at some length why private languages are held to be impossible [Wittgenstein (1999)]. [25]

Another line of attack on cognitivism is this. Supposing that there are sub-personal cognitive processes, *who* is doing this processing? Not the person in whose mind they occur: for, by hypothesis, they *underlie* his cognitive performances. (Surely *you* do not perform the computations that supposedly mediate between the disturbance of your retinae by various photons and you're your subsequent visual perception. These computations are done *in* your mind, but not by *you*: they are *sub*-personal.) There cannot be cognition without a cognizers. So the cognizers is some 'homunculus'. But then *his* cognitive performances have to be explained. If we are cognitivists about *his* mental activity, then a vicious regress threatens. On the other hand,

if we are not cognitivists about his mental activity, then we are conceding that we don't need cognitivism to explain mental activity. (See Appendix 1.)

Yet another attack on cognitivism is this: the very idea of sub-*conscious* — let alone sub-*personal* ('deeply unconscious') — processes is absurd. It is a *contradictio in adjecto* to speak of non-conscious mental phenomena; the essence of mentality is consciousness. The kinds of processes posited by cognitive scientists are sub-conscious. Indeed they are *deeply* unconscious: unlike the contents of the Freudian unconscious, they cannot be made conscious. So the very idea of such processes is an absurd one.[26] (See Appendix 2.)

For reasons of space, I will only be able to provide a substantive response[JK2] the first of these arguments (though I briefly deal with each of these arguments in the appendices)[JK3]. As I will try to show, one error in it is the idea that the sub-personal cognition posited by cognitive scientists necessarily involves a private *language*. The existence of a language, I will show, presupposes that users of that language have some *independent* access to the meanings borne by at least some of the expressions of that language. So if indeed there is sub-personal cognition — and I will try to show that there must be — it is not mediated through a *language*. Therefore, even if (what is doubtful) there can't be a private language, that has no bearing on whether cognitivism is correct or not.[27] [28]

II. WITTGENSTEINIAN PRELIMINARIES

Years before Brandom, Wittgenstein held (i) that what we call 'grasping' propositions really consists in knowing a language[29] and (ii) that one must be a member of a community to know a language. Brandom's reasons for holding (i) and (ii) draw heavily on Wittgenstein's reasons for holding them. So let me set forth Wittgenstein's reasons before I set forth Brandom's.

Wittgenstein's reasons for holding both of these views are to be found in his so-called 'Rule Following' and 'Private Language' arguments.

The Rule Following Argument

Suppose that you grasp some rule; suppose, for example, that you grasp the rule that is expressed by the symbol 'Fx=2x'. Let R be this rule. Obviously if you don't know how to apply R to particular values of x, then you can't be said to truly grasp R. If you don't know, or cannot readily figure out, that F(3)=6, then surely you don't really grasp R. So grasping R consists in, or at least involves, knowing how to apply it to particular cases.

How are you to know how to apply R? How are you to know, e.g., what F(37) is? Unquestionably there are correct and incorrect ways of applying R. (You are applying R correctly if you conclude that F(37)=74; incorrectly if you conclude that F(37)=80.) Let us suppose for a moment that there is such a thing as 'grasping R in a flash', as 'beholding' or 'intuiting' R; and let us suppose that you have managed to do this. In other words, let us assume, for argument's sake, that Plato was right: we grasp propositions (and *a fortiori* rules) through some analogue of perception. Surely, in intuiting R, you did not behold absolutely every instance of this rule; you did not behold that F(980)=1960. In any case, even if you did behold this particular application of R, there were others that you didn't behold. So you didn't behold every ordered pair constituting this function. Most, if not all, of these pairs were given to you only implicitly, i.e. they were given to you only in the sense that you grasped some rule that paired off the appropriate x's with the appropriate y's. So you do not know that F(980)=1960 merely in virtue of knowing R. Knowing R only enables you to *figure this out*. What would it be to figure this out? It would be to *apply* R to some particular case; and this, in turn, would be to *interpret* R. By itself, your knowledge of R is inert: *you* must

figure out that F(980)=1960; you must interpret R; you must bring your knowledge of R to bear on some particular case. Rules are like maps: maps don't tell you where to go: *you* figure out where to go by consulting a map. Similarly, *you* figure out that F(980)=1960 by consulting R.

But not just any interpretation of R will do; some interpretations are right and others are wrong. You must interpret R in some principled way; you must *follow some kind of a rule* in order to interpret R correctly. Let R_2 be such a rule. What we said about R is true of R2. R2 tells you how to apply R1 to each of infinitely many situations. Obviously R2 doesn't explicitly concern each of these situations: it only implicitly concerns them: so you must *figure out* how to apply R2 to each of these situations. Thus, you must know some *rule* that tells you how to apply R2. So one must grasp some rule R3 to know to apply R2. But what we said about R2 is true of R3. It is clear that we have launched ourselves on an infinite regress [Wittgenstein (1999), sections 139-241].

What is this supposed to tell us? As I see it, there are two lessons that we can extract from this argument. (It isn't clear to me whether these lessons are incompatible or, on the contrary, whether they are actually identical: the same message in two different forms.) The one message is this: the idea that we *grasp* norms — that we behold them in some way — leads to a vicious regress: we must grasp infinitely many norms to grasp a single one. Therefore there is no such thing as grasping norms: we simply aren't aware of norms.

This lesson is, it seems, to me too radical. There is plainly *some* sense in which people are aware of norms; and Wittgenstein seems to concede this. So I think the right lesson to extract from the foregoing is this: we *do* grasp norms; but we don't grasp them through some 'intellectual intuition'; we don't grasp them through some analogue of perception. The traditional view is this: we behold physical objects through the five senses; and we behold

abstract ones through some sixth sense. Wittgenstein is trying to undermine this view: whatever it is to grasp norms, this grasping doesn't have the ternary (subject-verb-object) structure that sense-perception (supposedly) has. In fact, in Wittgenstein's view, grasping a norm is not a *cognitive* affair at all: the cognitivist account of norm-grasping, argues Wittgenstein, leads to a vicious regress.

What could it possibly mean to say that grasping a norm is not a 'cognitive' affair? For it seems self-contradictory to say that 'grasping' a norm is not a 'cognitive' affair. We can't really answer this question until we set forth Wittgenstein's positive account of what it is to grasp a norm: and once we do that, this thesis will start to seem a little less paradoxical. But in the meantime we can make a couple of clarificatory remarks. To have a sense-perception of an object is to have a mental 'representation' of that object in one's mind. (In any case, this is the traditional analysis: significantly, it has been vigorously criticised in recent years by, among other, Wittgensteinians.) A 'representation' of an object is some kind of relatively discrete mental entity: a kind of icon or 'inner' counter-part to the external object in question. Wittgenstein rejects the idea that this model can be applied to our grasp of norms. Grasping a norm doesn't consist in having some kind of internal icon corresponding to that norm. So given that grasping a norm doesn't consist in being in possession of some 'inner counterpart' of that norm, what *does* it consist in? By way of anticipation, Wittgenstein's answer is essentially this. Having an ability — e.g. knowing how to play tennis — doesn't consist in being in possession of some mental icon; it consists in having certain dispositions. Similarly, grasping a norm doesn't consist in having some mental icon that represents that norm; it consists in having certain dispositions. (Grasping a norm is more like being able to play tennis than it is like seeing a rock.) More specifically, grasping a norm in

being disposed to act and think in ways that are consistent with that norm.

So grasping a norm is ‘non-cognitive’ in the sense that such a grasping doesn’t strictly speaking consist in *being aware* of anything: at least not in any sense even remotely comparable to that in which we are aware of things through sense-perception. Being aware of something seems to consist in being in possession of some representation of that thing. We do not grasp norms by being in possession of representations of them. So our grasp of them is in some important sense non-cognitive.

What Wittgenstein says about the question ‘how do we grasp norms?’ is supposed to apply to the question ‘how do we grasp propositions?’ (Actually, Wittgenstein is unclear about why, in his view, what he says the grasping of norms applies to the grasping of propositions. But it is clear that he *does* think that what we says about the former applies to the latter. What follows in the next few paragraphs is a reconstruction of Wittgenstein’s thinking, not a direct exegesis of it). This is because a proposition is a special kind of norm: a proposition is a kind of standard against which reality is compared. The proposition *the dog at the cat* constitutes a kind of norm to which I compare existing reality: if the dog did not the eat cat, then reality falls short of the norm; if the dog has eaten the cat, then reality has conformed to the norm. So we can think of a proposition as a kind of rule that assigns a truth-value to each of several different situations; in effect, a proposition is a rule assigns a *false* to a situation in which the dog has not eaten the cat, and a *true* to a situation in which the dog has eaten the cat. Given that propositions are rules or norms, it follows that what Wittgenstein says about norms applies to propositions: since we don’t grasp norms, we don’t grasp propositions.

This last line of thought no doubt requires elucidation. To grasp a proposition is to know what would make that proposition be true and what would make it be false. So to grasp a proposition is, in effect, to know some

rule that assigns *truth* to some situations and *falsity* to others. The proposition *the dog ate the cat* is thus a rule that assigns truth to some situations and falsity to others. Let R be this rule. There are infinitely many qualitatively[JK4] different possible situations which make this proposition true, and infinitely many which make it false; there are infinitely many qualitatively[JK5] different possible states of affairs in which the dog has eaten the cat, and infinitely many in which the dog has not eaten the cat. Obviously R doesn't *explicitly* concern each of these situations individually: R is not some infinitely long conjunction; it is not of the form *situation S1 is assigned truth and S2 is assigned falsity and S3 is assigned truth…*So R deals only *implicitly* with each of these possible situations. So to grasp R — that is, to know which truth-value it assigns to different situations — one must figure out how to *apply* R. Of course, one must be able to apply it in a principled — that is, a rule-governed — way. So to know how to grasp the first rule, one must grasp some second rule R2. But then what we said about the first rule applies to R2. R2 tells me how to apply R to each of *infinitely* many different situations. Obviously R2 cannot deal with each of these situations individually; R2 is surely not of the form *R1 assigns truth to situation S1, and R2 assigns falsity to S2, and R2 assigns truth to S2…*R2 deals only implicitly with each of these cases. So to grasp R2 — that is, to know how it applies to each of these situations — one must know how to apply it. Of course, must be able to apply it in a principled — that is, a rule-governed — way. So one needs to grasp some rule R3 to know how to grasp R2. Of course, what we said about R2 applies to R1: we have launched ourselves on an infinite and vicious regress.

Again, there are two ways to interpret this line of thought. The first is this: the idea that we grasp propositions is incoherent, as it leads to an infinite and vicious regress: so there is no such thing as grasping propositions. But this

seems too radical, and it seems inconsistent with certain remarks Wittgenstein makes. So the right way to understand it is this: we *do* grasp propositions, but we don't grasp propositions through some analogue of sense-perception. In fact, the grasping of propositions is not a *cognitive* affair at all.

What could it possibly mean to say that grasping a proposition is not a cognitive affair? Once again, we cannot fully answer this question until we set forth Wittgenstein's account of what it is to grasp a proposition. But what we said a moment ago about norms applies here. Grasping a proposition does not, in Wittgenstein's view, consist in being *aware* of anything. Being aware of something involves being in possession of some representation of that thing. And grasping a proposition doesn't consist in having some inner representation of that proposition: it consists in having certain dispositions.

But before we can state the nature of these dispositions, we must address a problem that Wittgenstein's anti-cognitivism leads to. Traditionally, it is thought that to understand a sentence S is to *know* what proposition to assign to that sentence: it is, in effect, to know some Tarski-biconditional — 'snow is white' is true just in case *snow is white*. But this conception seems to presuppose just the kind of 'cognitivist' account of what it is to grasp a proposition that Wittgenstein rejects. For presumably if one isn't capable of *representing* a proposition to oneself — that is, if one is not able to have a proposition 'in' one's mind — then *a fortiori* one cannot represent a Tarski-biconditional to oneself. So what answer does Wittgenstein give to the question: What is it for someone to understand a sentence?

His answer is basically this: Someone understands a sentence S when he knows how to use it. 'But doesn't someone's knowing how to use a sentence correctly depend on his first grasping some proposition?' No, says Wittgenstein. What a verbal symbol means is ultimately determined by when people use it; it is only people's usage of a symbol that determines when it is

correct to use it and thus what it means. So to know how to use a symbol, it is sufficient to use it in a way that is consistent with ‘communal forms of life’, i.e. that is consistent with other people’s usage of it. Basically, to understand a verbal symbol, e.g. a sentence, is to know how use it in a way that is consistent with conventional usage of it: to know a language is to be co-ordinated with other people in respect of when one uses certain symbols. So knowing a language consists not in grasping various Tarski-biconditionals — consists not in knowing that ‘Bob is true’ means *Bob is tall* — but in being sensitized to other people — in being acculturated in a certain way.

‘But this account suffers from circularity’, one might argue. ‘Wittgenstein says that one knows what S means when, and only when, one knows how to use S. This analysis is supposed to show that for one to ‘grasp a proposition’ isn’t for one to have some kind of intuition of a mysterious, other-worldly being, but only to be sensitized to certain behaviours. But to know how other use a symbol S is itself to know a proposition. So Wittgenstein explains what it is to grasp a proposition in terms of itself.’

This objection fails to distinguish between knowing-that and knowing-how. To know how to hit a baseball or to ride a bike is not to be in touch with some proposition; it is simply to be competent in a certain way.[30] Knowledge-how is not cognitive, i.e. it doesn’t involve a grasp of propositions. One can know how to ride a bicycle without knowing any propositions. There are no ontological mysteries associated with knowing-how. For Wittgenstein ‘knowing what S means’ is a kind of knowing-how, not a kind of knowing-that. To know what S means is to be co-ordinated with other people in a certain way; it is to be competent to play a certain communal game. Knowing how to box doesn’t consist in grasping certain propositions: it consists in knowing-*how* to react to certain situations; it consists in a certain kind of sensitivity, not a certain kind of *cognitivity*.

Analogously, knowing how to use a symbol is to have a certain kind of sensitivity, not cognitivity. If Wittgenstein were explaining what it is to know what S means in terms of some other kind of propositional knowledge — i.e. in terms of knowing-that — then his analysis of what it is to grasp a proposition would be circular. But he doesn't; so it isn't. (For the record, my own view is that knowing-how *does* reduce to knowing-that — that knowing *how* play the piano or do math *does* consist in internally representing various rules and is thus comparable to knowledge of a fact. To be sure, where knowledge-how is concerned, the internal representation is housed in a much deeper stratum of mental activity than where knowledge-that is concerned. But, apart from this, there is no structural difference between knowing-how and knowing-that. So I will presently argue, in any case.)

Let us sum up. For Wittgenstein, to know what S means is to know how to use S; and to know how to use S is simply to co-ordinated with other people in a certain, pre-cognitive way. Grasping a proposition is not a matter of having certain intuitions, but only of having certain kinds of sensitivity to the symbol-producing behaviour of one's brethren.

If Wittgenstein's analysis is correct, it is inconceivable that one should grasp a proposition prior to knowing a language; for grasping a proposition consists solely in knowing a language.[31]

The Private Language Argument

Now we must ask: why does Wittgenstein hold that one must be a member of a *community* to speak a language? Couldn't one know a 'private language'? Couldn't one invent a language of one's own, and grasp propositions by knowing how to use sentences belonging to that language? Wittgenstein's answer is: no. His reason for holding this lies in his 'Private Language Argument.' What is this argument?

For an expression S to have meaning, there must be correct and incorrect

ways of using it. If it is correct to use S under *any* circumstances, then S will be infinitely ambiguous and therefore meaningless. This is the idea underlying the alleged impossibility of a private language: an expression belonging to a private language can be used in *any* way with impunity; therefore, in effect, any way of usage such an expression is right; and, consequently, any such expression is infinitely ambiguous and thus meaningless.

Suppose someone tries to invent a private language of his own. By a 'private' language I mean a language that is *not* simply a translation of some existing public language. Obviously I could invent a private language by translating words of English into words that I come up with. What I wish to consider now is: Could someone invent a language of his own that was *not* merely a translation, an isomorph, of some existing public language? Could someone who did not already have a word meaning *cat* undertake to assign this meaning to some sound-type? Let us suppose that someone could to see if any contradiction or absurdity follows. Let us suppose that someone (call him 'Fred') who did not already speak a language considered some sound — say 'blark' — and stipulated that 'blark' was to mean *cat*; and then considered some other sound — say 'snark' — and stipulated that this was to mean *dog*; and so on. Of course, Fred will have to do more than assign referents to certain sounds; he must also (e.g.) assign meaning to grammatical devices (an example of a 'grammatical device' would be adding s to a noun to pluralize it); he must also give meaning to non-lexical terms (terms like 'not', 'or', 'if'): basically, Fred will have to assign meaning to each of several primitive symbols, and he will also have to lay down recursive rules that specify the meanings of complex expressions of his language in terms of the meanings of these simple expressions.

Let us suppose Fred does all of this, and that he thereby creates a fairly

rich language for himself. Let L be this language. Let us suppose that Fred records his thoughts in a diary, and that the states these thoughts in L. If Fred uses a particular sentence S, how will know that he has used it correctly? To what standard will he appeal?

If a member of a linguistic community uses or interprets an expression incorrectly, there are consequences. If I turn right when it says ‘no right turn’, I will be given a ticket. If I ask for ‘chocolate’ when I really want vanilla, I will not get what I want. Basically, if I use an expression incorrectly there are negative consequences; and these consequences are imposed on me by other people. Fred uses L only to communicate with himself; so no one besides Fred will punish his incorrect uses of expressions of L. But whenever Fred uses some expression of L, he will always *think* that he is using it correctly. He has only his own opinion as to whether he is using such an expression correctly. As a matter of logic, Fred can’t really consider his own opinion to be wrong. (If you consider some opinion to be wrong, then it isn’t *your* opinion, even if it used to be.)

Of course, there will be some fact of the matter. Fred’s original definitions may be such that, according to those definitions, S means ‘I saw a cat today’, even though Fred currently takes S to mean ‘I killed a dog today.’ But of what consequence would these original definitions be? These definitions form a part of L only in so far as they have some effect on how expressions belonging to L are used; and they have such an effect only through the medium of Fred’s *opinion* as to what these original definitions were. Fred’s opinion will be no less efficacious when it is dead wrong than when it is right. So these definitions would be idle. (Even though they might have been a part of Fred’s language in the past, they are no longer such.) Only Fred’s opinions will be effective. Consequently, expressions of L will mean whatever Fred *thinks* they mean. Since Fred is the only user of L, and

since Fred will always think that he is using a given expression correctly, it follows that expressions of L cannot be used incorrectly. This in turn means that these expressions will be completely ambiguous and therefore meaningless. So a private language is an absurdity. Only a creature that belongs to a community can speak a language: for only a *public* rule has the stability — the imperviousness to this or that person's mistaken opinions — characteristic of a *true* rule [Wittgenstein (1999), pp. 243-71].

Brandom's View

Why does Brandom hold (i) that we must know a language to grasp propositions and (ii) that we must be members of a community to know a language? His reasons are close to Wittgenstein's. The individual cannot privately grasp rules or therefore propositions, as the rule-following and private language arguments supposedly show [Brandom (1998), pp. 18-22]. Norms are embodied in communal linguistic practices: it is here that they are initially 'made real' [Brandom (1998), pp. 34-46]. These norms cannot be embodied in the practices of a 'private linguist', as Wittgenstein, to Brandom's mind, has shown. So for the individual to grasp norms and therefore propositions, it is *necessary* that the individual enter into these practices [Brandom (1998), p. 53].

At this point Brandom goes a little bit beyond Wittgenstein. It is certainly *necessary* to be in step with others, in a certain way, to know a language and, therewith, to grasp propositions. But is it *sufficient*? Care must be taken to avoid the fallacy of *regularism*, the view that a creature is *really* governed by some kind of norm so long as it is acting *as if* it were governed by that norm. A person can be 'in step' with others, at least as far as his overt behaviour is concerned, without thereby being governed by any norms: he could just be a zombie who, without being conscious of anything (let alone norms), speaks

and acts *as though* he grasped the relevant norms. So falling into step with others is not enough to convert a creature that is oblivious to norms into one that is governed by them [Brandom (1998), pp. 40-41]. What else is required?

According to Brandom, norms are 'implicit' in the activity of certain social groups, and people that are not yet governed by norms come to be governed by norms by entering into these groups. How does this happen exactly? The group rewards the individual for following the relevant norms and punishes him for disobeying them. (The group does not always *know* that this is what it is doing: after all, the group has only an 'implicit' awareness of its own norms.) As a result of this conditioning, the individual assimilates the norms that are embodied in the group's behaviour; these norms become *his* norms [Brandom (1998), pp. 42-46].

Why, in Brandom's view, are norms initially *implicit* rather than explicit in the group-psyche? Brandom thinks that the rule-following argument is sound and that, consequently, were one to suppose an *explicit* awareness of norms to be antecedent to an implicit awareness of them, this would lead to a vicious regress: an explicit awareness of a rule (or, more generally, a proposition) R_1 needs to be supplemented by some awareness of some rule R_2 — otherwise one won't know how to apply R_1 — and an explicit knowledge of R_2 needs, in its turn, to be supplemented by a knowledge of some rule R_3, and so on ad infinitum [Brandom (1998), pp. 18-21, also pp. 26-30].

With regard to the doctrine that one must grasp rules *explicitly* to be governed by them, Brandom refers to this doctrine *regulism*. The rule-following argument, thinks Brandom, shows that regulism is false [Brandom (1998), pp. 18-21]. But at the same time, as we have already remarked, in order for one's thought and behaviour to be norm-governed, one must have *some* kind of awareness of rules. So, reasons Brandom, this awareness must

be *implicit*, not explicit. But as we know, according to Brandom, one cannot have such an awareness privately: one must be a member of a community. So, Brandom concludes, it is ultimately *communities* that have this awareness (individuals have it only by virtue of their participating in communal practices); and, once again, this awareness must, initially at least, be implicit. In a word, awareness of norms is 'implicit' in a community's practices (in particular, in its linguistic practices); and the individual acquires such awareness by becoming a member of a community — in particular, by participating in its linguistic practices [Brandom (1998), pp. 37-46].

Let us sum up. What is Brandom's answer to the question: how do individuals come to be aware of norms? It is this. They enter into groups that, collectively, have an implicit awareness of these norms. The individual is punished for behaving in ways proscribed by the relevant norms, and rewarded for behaving in ways mandated by them; and this somehow makes him aware of these norms.

III. CRITICISMS

Both Wittgenstein and Brandom agree that it is not possible to grasp meanings or rules *prior* to knowing some language. This position, I will now try to show, leads to an untenable analysis of linguistic competence and, indeed, of every type of cognitive competence.

Suppose that, by all appearances, Joe is a good mathematician; he consistently solves hard problems. A cognitive psychologist, like Chomsky, would explain this by saying that Joe has an 'implicit grasp' of some set of rules or principles which he brings to bear on particular problems. (In this, cognitive psychology would probably be in agreement with common sense, so far as the latter has any opinion on the matter.)

But this kind of explanation would not be satisfactory to everyone. Some would regard it as defective on methodological grounds: it explains the

obscure by the more obscure; a certain kind of competence is ‘explained’ by positing an utterly mysterious ‘inner condition.’ This is, we are told, like explaining the soporific properties of opium by ascribing a ‘dormitive virtue’ to it.

As we have seen, Wittgenstein has a more specific reason for deploring this kind of explanation: no mental state or condition can possibly constitute a grasp of a rule; no mental state can have such scope. Brandom accepts Wittgenstein’s view. So Brandom and Wittgenstein identify the grasping of propositions with the having of dispositions. They reject the ‘perceptual model’ of what it is to grasp a proposition; they reject the idea that grasping a proposition consists in having some internal representation of that proposition. This leads them to identify the grasping of a proposition with the having of dispositions.

But how are these philosophers to account for the fact that Joe consistently solves difficult problems? In what, according to them, does his mathematical competence consist? They reject the idea that Joe’s mathematical competence consists in his grasping certain truths. So in what is that competence supposed to consist? There is only one line they can take: Joe’s competence consists solely in his having certain *dispositions* to act a certain way or to have certain mental states. Joe is *disposed* to make certain noises or to have certain awareness; and his understanding consists entirely in these dispositions.

Of course, by hypothesis, these dispositions do not spring from any underlying grasp of rules: if one says that a disposition *springs from* a grasp of a proposition, one is barred from *identifying* the disposition with the grasping of the proposition. The dispositions are cognitive rock-bottom. To be a good pianist, a good poet, a good mathematician — to have any kind of cognitive competence — is to have certain dispositions, and that the

dispositions in question do not rest on top of some kind of cognizance of rules: the dispositions *are* the cognizance in question.

This last qualification is crucial. The purpose of dispositionalist analyses is to *eliminate* the need to hypothesize some kind of 'inner representation' of rules or propositions. So a dispositionalist analysis cannot say that the dispositions in question exist because of such a representation. In so far as one has any kind of representation of a proposition, the dispositions *are* that representation. The dispositions *are* one's cognizance of the proposition.

I should clarify an important component of the foregoing line of thought. Wittgenstein and Brandom make the inference from 'grasping a proposition (or, more generally, a norm) doesn't consist in having an internal representation of that norm' to 'grasping a proposition (or rule) consists entirely in the having of dispositions of some kind.' What is the warrant for this inference? Any kind of cognitive competence expresses itself entirely in various productions — in various behaviours and mental contents. Suppose we reject the idea that these various productions emanate from some *understanding* — that is, from a grasp of propositions. In that case, we are obviously barred from explaining these productions in terms of the operation of *reasons* and *motivations.* For these reasons and motivations would inevitably incorporate just the kind of 'cognitive' grasp of propositions that Wittgenstein and Brandom believe not to be possible: these reasons and motivations would inevitably be such that it would not be possible to conceive of them except *as* just being, or at least involving, representations of propositions. An illustration may help. Joe is asked 'if Fred has two apples, and Mary has two apples, then how many apples to Fred and Mary have collectively?'. Joe says '4 apples.' Obviously *one* explanation of Joe's conduct is this: Joe internally represents some arithmetical proposition and, on the basis of the content of that representation, is motivated to produce

what he perceives to be the correct response. Obviously Joe's response betokens some kind of disposition: a disposition to respond with certain outputs to certain inputs. This is uncontroversial. *But* this disposition is *grounded* in some representation of a proposition: the disposition is not cognitive rock-bottom. No — what is cognitive rock-bottom is some kind of internal representation of a proposition; and the disposition emanates from this disposition. This is the view *I* wish to defend; and it is the traditional view.

But Wittgenstein and Brandom cannot take this view. So they have set themselves a difficult task: they must explain *why* (e.g.) Fred says 'four' after hearing 'what is 2+2?' Now unquestionably Fred's responding with that output to that input expresses some kind of disposition: that is undeniable. In general, any performance that betokens cognitive competence obviously corresponds to some kind of disposition. In virtue being a competent pianist, one has a disposition to play certain notes at certain times. In virtue of being a competent mathematician, one has a disposition to give certain answers to certain questions. This is not controversial. What is controversial is whether these dispositions are *grounded* in internal representations of rules of some kind: what is controversial is whether these dispositions radiate from some inner grasp of a proposition. As we've seen, Wittgenstein and Brandom say: No — this disposition doesn't radiate from an internal representation of a proposition. Given that it doesn't radiate from an internal representation of a proposition, it would seem to follow that it doesn't radiate from *any* internal representation. What kind of representation could possibly do the job in question *besides* a representation of a proposition? What kind of representation *other* than a representation of some general rule could possibly ground or validate the multifarious dispositions associated with, say, musical or linguistic or mathematical competence. Surely none. Indeed, it seems that

no kind of *mental content* — whether a representation or not — other than a representation of a proposition could ground the various dispositions associated with a given type of cognitive competence. So *if anything* mental *grounds* the dispositions associated with a kind of cognitive competence, it must surely be a representation of some general proposition or rule. So, since Wittgenstein and Brandom *reject* the idea that there are such representations, they are forced to see the dispositions in question as *ungrounded*: as cognitive rock-bottom. So for Wittgenstein and Brandom to have a kind of cognitive competence is not to have dispositions that are *grounded* in a grasp of propositions or rules: it is *just* to have disposition; it is to have dispositions that are not in any way *grounded*. The dispositions are rock-bottom. To sum up, given their anti-Platonist epistemology, Wittgenstein and Brandom must take a dispositionalist view of (e.g.) mathematical competence; and this is just the view that they do take.

But is such a 'dispositionalism' about cognitive competence tenable? No. There are a couple of different reasons for this. The first reason must be set forth in stages. To begin with, everyone — including Brandom and Wittgenstein — must concede that *if* Joe's saying '4' is to betoken any kind of cognitive competence at all, then it must have *some* mental cause. If a creature's performances are the result of purely physical dispositions (by 'purely physical' I mean 'lacking a mental dimension altogether'), then those performances would be the work of a robot. Even hard-line materialists hold that physical processes must have a mental dimension *if they are to qualify as cognitive.*

To be sure, materialists hold that a physical process qualifies as mental wholly in virtue of its physical characteristics. But every materialist holds that if a physical process does not qualify as mental, then it is not *cognitive*: it does not embody *understanding* of any kind. So some kind of *mental* activity

must mediate between Joe's being stimulated in a certain way (his seeing, e.g., the following inscription on a piece of paper: 2+2=?) and his subsequently acting a certain way (his writing 4): even Wittgenstein and Brandom will concede this.

But — this takes us to the next stage of the argument — not just any kind of mental mediation will do: the mental states intervening between stimulus and response must be such that, *by virtue of having those mental states,* Joe's response embodies some kind of *understanding* of the stimulus.

Suppose that after seeing the inscription 2+2=? Joe has a variety of mental states: he is joyful, then melancholy, then he feels a tickle, then he has an itch; let us also suppose that he has no mental states *besides* these. After having these mental states, Joe ends up writing 4; it may even have been one of these states that *caused* Joe to write this. To be sure, Joe is not a complete robot; after all, robots don't feel joy or melancholy. But Joe's performance was still robotic because, although it had a mental cause, it didn't have the right kind of mental cause. A performance that has a mental cause is not, on that account alone, cognitive or non-robotic. (By 'cognitive' I mean 'embodying understanding or a grasp of a proposition.') If Joe has been trained by a hypnotist to write 4 every time he sees 2+2=?, then Joe cannot be said to have solved any problem if, on some particular occasion, he writes 4 after seeing 2+2=?. So what kinds of mental states of affairs must mediate between stimulus and response — between seeing 2+2=? and writing 4 — if Joe is to be said to have solved a problem of mathematics? The answer is obvious: some kind of *understanding* must intervene. In so far as no understanding intervenes, Joe's response is 'robotic.' But if we say that understanding mediates between Joe's seeing 2+2=? and writing 4 then we have given up on dispositionalism. Joe has the right response because he understands the input he is given. So Joe's disposition to have the right

response is *grounded* in understanding. Thus understanding *grounds the disposition*; the disposition does not ground the understanding. Put another way, a disposition that is not grounded in understanding is 'robotic'. But once we concede that a disposition must *presuppose* the presence of understanding if that disposition is to qualify as a kind of cognitive competence, then we are obviously giving up on the idea that dispositions ground understanding — we are giving up on the idea that to have understanding, to have a kind of cognitive competence, just is to have a disposition of some sort.

There is a second reason to reject dispositionalism, at least as a general explanation of cognitive competence. A good mathematician, like Joe, is able to solve an indefinitely extensible class of problems. Ascribing one disposition to Joe won't even come close to accounting for his mathematical performances. If my knee is struck a certain way, my leg jerks: *that* is an example of a disposition. So supposing that we are dispositionalists about cognitive competence, then we must say that Joe's mathematical competence comprises, not just one, but several dispositions. But will any finite number of dispositions suffice? A good mathematician is always able to assimilate new principles and techniques. There isn't any pre-set limit to the variety of such principles and techniques. So we would have to ascribe an infinite number of different dispositions to any good mathematician. Actually, we'd have to do the same for *anyone* who had mastered so much as grade school arithmetic: for no two mathematical problems, no matter rudimentary, are quite alike. (If they were completely alike, then there would only be one problem, not two.) So one disposition can enable Joe to solve, at most, one problem. So we end up having to postulate an infinite number of different dispositions to Joe. But how do we explain the fact that Joe just *happens* to have all these dispositions? If we account for this by saying that all these dispositions flow from some underlying grasp of mathematical principles,

then their existence is totally intelligible. But if we deny that there is any such underlying representation, then every time we postulate a disposition to explain how Joe solves problem, this postulate is completely *ad hoc.*

Any kind of cognitive competence is associated with *infinitely* many different dispositions. The disposition associated with one's saying '5' in response to hearing 'what is 2+3' is obviously not identical with the disposition associated with one's saying '4' in response to 'what is 2+2'. But just as obviously these dispositions are *connected* with each other. Now if we say that the definitions *are* one's mathematical competence, then we are saying that these various dispositions are not grounded in anything and, therefore, that they are not grounded in the *same* thing: so if we are dispositionalists about e.g. mathematical competence, then we end up saying that the various dispositions associated with such competence are not connected to each other: Joe's disposition to say '4' after hearing 'what is 2+2' becomes unrelated to his disposition to say '5' after hearing 'what is 2+3'. But it is patently obvious that these dispositions are connected: what connects them is that they emanate from some common source: some underlying *understanding* binds them. Since the understanding *underlies* them, it cannot be *identical* with them.

The basic problem with the dispositionalist view is that it tries to assimilate the normative to the causal order. Joe's performances can, without artifice, be mapped onto some normative order. If we ascribe an understanding of this order to Joe, then this isomorphism is explicable. But if we deny that Joe has any such understanding, then this isomorphism is pure happenstance. But it obviously isn't pure happenstance.

Given that we cannot 'dispositionalize' mathematical competence, how *are* we to explain it? It is true that mathematical competence *comprises* an infinite number of dispositions. It is also true, as we've just seen, that

mathematical competence doesn't *consist* of such dispositions. For these dispositions must have a common ground; and if they *are* the mathematical competence, there will be no such ground. Clearly what we must say is that mathematical competence consist in internal representations of mathematical principles, and that it these internal representations that provide the common ground for, and source of, the aforementioned multiplicity of dispositions.

IV. THESE CRITICISMS EXTENDED

Now I would like to show that what we have said about cognitive competence generally applies to linguistic competence specifically.

On Wittgenstein's view, knowing a language does not consist, even in part, in knowing rules; for according to Wittgenstein, one must know a language in order to grasp rules. (For Wittgenstein, we know rules *by way of knowing language*. The traditional view — the Chomsky/*Tractatus* view — is that we know language *by way of knowing rules.* I am trying to defend the latter view.) But knowing a language certainly *seems* to presuppose a knowledge various rules: it *seems* to consist in, or at least involve, knowing rules like: 'snow is white' means *snow is white.* We easily understand sentences that they have never encountered before. The obvious explanation of this is to say that we have internalized some set of rules saying how sentence-meanings result from word-meanings: that we have internal representations of some rules and that these representations guide us in assigning meaning to sentences that we haven't heard before. Michael Dummett puts this point very well:

> The fact that anyone who has a mastery of any given language is able to understand an infinity of sentences of that language, an infinity which is, of course, principally composed of sentences which has never heard before...can hardly be explained otherwise than by supposing that each speaker has an implicit grasp of a number of general principles governing the use in sentences of words of the language [Dummett (1978)].

Wittgenstein-Brandom cannot countenance this point of view. So how would Wittgenstein-Brandom account for our being able to understand sentences we have never encountered before? He must say that we are simply *disposed* to have certain reactions to them; that we are disposed to have certain mental and behavioural reactions to them: it is just some structural property of our minds or brains — some property that doesn't embody cognizance of norms — that induces us to react with the right sentences to new situations and with the right mental contents to new sentences.

This is surely untenable. If we haven't grasped the relevant rules, then how is it that we so frequently have the *right* reactions to new sentences? What would account for our dispositions' so often being the *right* ones unless these dispositions were themselves the results of some deeper awareness of norms? It seems that, unless we were aware of the relevant norms or rules, we wouldn't have the right dispositions.

> You frequently talk of having the right reactions to sentences, and of producing the right sentence under a given set of circumstances (i.e. of having the right linguistic reaction to a given set of circumstances). But right relative to what? Your argument against a dispositionalist analysis of mathematical competence made some sense because there is a body of objective mathematical truths: consequently, if a dispositionalist analysis of our mathematical competence were correct, then it would be a miracle that our dispositions so frequently resulted in our having the right beliefs about this domain. But is there any domain of grammatical truths? Surely not. If everyone believe that the twice two were five, this belief would not, on that account, be correct. But if everyone regarded as grammatical the sentence 'John Mary loved' (meaning: John loved Mary) then this sentence would, on this account, be correct. So your criticisms of the dispositionalist view of mathematical competence do not transfer over to

a viable criticism of linguistic competence.

This objection ultimately doesn't carry much weight. It is true that, if everyone said 'John Mary loved' instead of 'John loved Mary' then the former would be grammatical while the latter wouldn't be. But grammaticality is a property of an infinite number of sentences of any given language; linguistic behaviour can *directly* result in only a finite number of sentences' being grammatical. Linguistic behaviour does indeed result in an infinite number of sentences' being grammatical. But it does this only *in*directly, i.e. it does this through the mediation of rules — rules that are projected somehow from those linguistic productions whose correctness is *directly* determined by human conduct. These rules constitute the standard against which it is to be determined whether a person's linguistic productions are appropriate to a given situation and whether a person's intellectual productions are appropriate given a particular linguistic stimulus. So it does make sense to talk of right and wrong linguistic productions, just as much as it does to talk of right and wrong mathematical productions.

> But mathematical rules aren't arbitrary. The commutative law for addition is a fact. By contrast, grammatical rules are arbitrary. How can it make sense to speak of right and wrong relative to an arbitrary set of rules?

It is true that in mathematics the rules are not arbitrary, whereas in grammar they are. But this is irrelevant. Given that certain grammatical rules are in place, it makes perfectly good sense to speak of right and wrong linguistic productions. It is a hard fact that 2+2=4 and that 2+2≠5. It is just as hard a fact that, given the rules of English grammar, the sentence 'if Smith weren't so parsimonious, he would enjoy life more' is grammatically correct.

> Consider the class of expressions that have actually been produced. There is an infinite number of different grammars that can be projected from

these expressions. So one person's linguistic productions are no more right or wrong than the next person's. If Joe says 'me no love Mary', he cannot be said to be wrong, only deviant (in a statistical, not a normative, sense). Similarly, if Joe takes 'I love Mary' to mean 'I want to have a blue suit', he can only be said to have extrapolated in a non-standard way from a give set of data. So there is no right and no wrong where linguistic productions are concerned.

This objection virtually answers itself. It is true, as the objector points out, that the rules that we project from linguistic behaviours are arbitrary; different rules could be projected from them. There is no reason why, given how people's actual linguistic behaviour, a grammar could be projected, from this behaviour, whereby 'I love Mary' meant *I want to have a blue suit.* But this isn't the projection that is in fact operative; certain rules are projected and not others. Given these rules, it is a hard fact that 'I love Mary' means *I love Mary* and not *I want a blue suit.* Given these rules, certain linguistic and intellectual productions are right and others are wrong. The fact that some other set of rules, some other grammar, could have been projected is irrelevant.

It is a fact that a certain grammar has been countenanced. When we talk about rightness and wrongness in people's linguistic productions, we mean rightness and wrongness relative to this grammar. Consider the sentence 'if Alfred were more easy-going, and if he had little more money, he would be capable of enjoying a quick trip to Naxos.' I know that this sentence is perfectly grammatical. I don't have to go out and examine people to determine this. I know right now that every English person I might query (with some obvious and irrelevant exceptions — e.g. people with certain cognitive defects) would agree with me that this sentence was grammatical. I know this because I know that all these people countenance one grammar and

not some different one. There is no other way that I could know this. If people were not law-like in their linguistic productions and in their intellectual responses to linguistic productions, then I'd have no way of knowing whether they would regard a given sentence as grammatical or not. But you and I are both unerring in our judgement as to what people of our dialect of English will regard as grammatical, even where sentences that no one has ever heard before are concerned; and this is possible only because people are law-like in their linguistic behaviour. Straight induction off of people's linguistic behaviour wouldn't enable us to do this (by 'straight' induction I mean induction that didn't involve imputing a cognizance of rules to people); we can predict people's linguistic behaviour only by imputing to them cognizance of a certain rules and by ascribing causal efficacy to this cognizance. The dispositionalist account of linguistic competence cannot account for this law-likeness, this predictability, in people's linguistic behaviour.

Basically, there is a grammar that people *in fact* conform to, whether cognizance of this grammar is causally efficacious in people's conduct or not. (This grammar is not all that hard to state, at least in rough outline; every elementary school grammar book does a passably good job.) It would be an utter miracle if (i) people's linguistic behaviour were such that it could be described *as though* people were following these rules, even though (ii) they were *not* in fact in following these rules (or at least some rules isomorphic thereto). Basically, in order to explain the isomorphism between the normative and the actual order — between what people do and what, given a certain grammar, they *ought* to do — it is necessary to suppose that some kind of cognizance of the normative order (in this case a grammar) is responsible for people's productions. What we said about dispositionalism as an analysis of mathematical competence applies to dispositionalism as an

analysis of linguistic competence.

Also, to echo a point made earlier, if my ability to speak English *consisted* in my having various dispositions that weren't bound together by some kind of internal representations of rules, then these various dispositions would have no connection to each other: not being bound together by anything, it would be pure happenstance that they all co-existed. But they are not just fortuitously co-existence. I have a disposition to regard instances of the sentence-type 'the dog ate the cat' as well formed, and also to regard instances of the sentence-type 'the mouse ate the tiger as well formed. These are obviously different dispositions. But just as obviously, they are connected. Now if my grasp of English were *identical* with my having so many dispositions, then nothing would *underlie* the two dispositions just mentioned: for those two dispositions would be cognitive rock bottom. Consequently those two dispositions would be unconnected to each other: indeed given *any* two of the dispositions associated with my grasp of English, they would be unconnected to each other, if we accept the dispositionalism of Wittgenstein and Brandom. But obviously these various dispositions are connected. Now what could possibly connect them — what could possibly provide a common ground for them — if not internal representations of the propositions that validate them?

V. WITTGENSTEINIAN-BRANDOMIAN COUNTER-OBJECTIONS[JK6]

If knowledge of rules doesn't underlie, and thus pre-exist, knowledge of language, then knowledge of language consists entirely in having a lot of dispositions. But having a lot of dispositions doesn't add up to having a grasp of meanings or, therefore, to being able to understand sentences. Pavlov's dog doesn't *understand* the ringing of the bell; it just reacts to it. *Reacting* to the sound 'fire' isn't understanding it. Reactions that are not mediated through a grasp of concepts are not cognitive; they don't betoken

understanding.

In response to this, Wittgenstein could say that I have the situation backwards. What really happens is that we are *first* disposed to have certain reactions to certain sounds, and to react with certain sounds to certain situations, and we *then* elevate these *de facto* dispositions to the status of norms. In other words, there are first various mental and physical behaviours that embody no awareness of norms. But later on these behaviours are taken to embody norms. (For charity's sake, we won't ask how it is that normless creatures manage to convert some pre-normative practice into a norm-governed one. Surely a grasp of norms is required to pull this off.) A comparison might be illuminating. It is now *de rigueur* to buy presents for a friend on his birthday: one is violating a kind of law if one fails to do this. But it is quite conceivable that, a long time ago, people bought presents for their friends on their birthdays out of a spontaneous desire to them happy. So a behaviour that didn't embody awareness of norms came to do so later on. Similarly, a long time ago people might simply have reacted to the presence of fire by blurting out 'fire'. Later on, this behaviour was taken to be the *appropriate* reaction to fire.

This line of thought could, at most, explain why words like 'fire' mean what they do (though I will question this in a moment). But it definitely couldn't explain why expressions like:

'if Mr. Smith were to see this in his current condition, he would probably die of a heart attack, assuming he didn't have his heart pills with him'

mean what they do. This expression has probably never been used before. So it is not only people's usage of this expression that gives it its meaning. Of course, people would be disposed to use this sentence in response to certain kinds of situation. (It would be easy to think of a situation where you might be inclined to use this expression.)

Of course, it would certainly be correct to say that this sentence means what it does because of how people use the expressions occurring in it, even though no one has ever used that expression, as a whole, itself. But if that is true, then there must be *rules* saying how the meaning of complex expressions (like ‘if Mr. Smith saw this...’) results from combinations of simple expressions. Unless the meaning of ‘if Mr. Smith saw this...’ is to be fixed *purely* by convention — unless it means what it does because of when people happen to use *that exact expression*, taken as a whole — we are led back to the position that ability to understand certain expressions, namely complex ones, involves an implicit grasp of rules. This, of course, is just what Wittgenstein-Brandom cannot say.

A die-hard Wittgensteinian-Brandomian might contest this last point. He might say that this sentence (‘if Mr. Smith...’) means what it does because of when we *would be* disposed to use it. But people would be disposed to use this sentence only because they already knew a language; i.e. because they were already aware of some set of rules or norms and, in light of this awareness, they regarded a certain sentence as being appropriate to the circumstances. In general, people are disposed to use complex expressions when they do only because they *already* know a language. Most of our sound-producing behaviour *presupposes* a knowledge of language. So language is not just a lot of initially non-norm-governed sound-producing behaviour that becomes, in some mysterious way, norm-governed sound-producing behaviour. In particular, in so far as knowing a language involves knowing how to use complex expressions, then knowing a language is not reducible to having certain dispositions. So language cannot result *entirely* from pre-cognitive reactions. (By a ‘pre-cognitive’ reaction, I mean one that doesn’t result from awareness of norms.)

A Wittgensteinian-Brandomian might disagree with this last point; he

might argue as follows. The *basis* of any language is just a lot of sound-producing behaviour elevated (in some way or other — we won't ask how) to the status of norm-governed behaviour. (People used to say 'fire' when there was a fire nearby. At first they just did this: it wasn't right or wrong. But then it came to be *de rigueur* to say this when, and only when, there was a fire nearby.) Although it is implausible to say that

> 'Mr Smith would die of a heart-attack if he saw this, provided he didn't have his part pills with him'

means what it does because people just blurt it out in certain situations, it is not implausible to assume that 'fire' means what it does because people just blurted it out in certain situations (though, by way of anticipation, we will see that this is false). Moreover, it is reasonable to assume that people have certain set reactions to *combinations* of blurtings. For example, people might just have a disposition to react a certain way to 'fire and bear' or to 'no fire' or to 'if no fire then no hot meal'. These complex blurtings would, in some or way or other, be promoted to the level of norm-governed sound-productions, just like their simple counterparts: so just as 'fire' went from being a sound that people simply *happened* to produce to one that it was *correct* to produce; so 'fire and bear' went from being a sound that people simply *would* be disposed to produce to being one that it would be *right* to produce. If people had dispositions of this second class, then we could account for the meaningfulness of complex expressions *without* supposing that anyone had an implicit grasp of any rules: it would just be 'dispositions all the way down.'

Is this tenable? No. It isn't hard to see the problem with this analysis: we end up having to hypothesize an infinite hierarchy of dispositions, without these dispositions' having any *raison d'être* other than that their existence enables us to hold on to some essentially *a priori* belief that it must be

possible to explain linguistic competence without assuming that people have a pre-linguistic grasp of rules. To explain why people understand ‘bear and fire’, we have to hypothesize that people just happen to have a disposition to respond to ‘bear and fire’ the right way, even though they’ve never encountered *this* exact expression; and to explain why the end up why people respond the right way to ‘if bear and fire and spear, then bear-meat for dinner’ the right way, we end up having to posit *another* disposition; and so on, for each complex expression. If we assume, as Chomsky and Dummett do, that linguistic competence involves a pre-linguistic grasp of rules, then we have no trouble accounting for the existence of such dispositions: we have the right reactions because we grasp propositions and therefore know what the right reactions are. But if don’t make this assumption, then we have to impute an infinite number of dispositions to people, without having justification for so doing.

Let us sum up. There is no way to account for our being able to understand produce new sentence without assuming that we have some kind of pre-linguistic grasp of rules. So linguistic competence doesn’t just consist in having a lot of dispositions; for we couldn’t have the right dispositions without having a grasp of rules. This means that Wittgenstein’s rule following argument is not sound; for this argument is supposed to show that ability to follow rules is the product, not the cause, of the ability to speak a language.

VI. ANOTHER CRITICISM OF WITTGENSTEIN-BRANDOM

If Wittgenstein-Brandom is right, then to understand an expression is to be competent to use it in accordance with convention; and to be competent to use it in accordance with convention is to have certain dispositions: it is to react in certain ways to certain stimuli. (Wittgenstein-Brandom cannot say that to be competent to use an expression consists in grasping rules or

meanings. He would incur a vicious regress if he said this. So he must say such competence consists in having certain blind dispositions.) But does having a lot of dispositions add up to *understanding*? What distinguishes us from Pavlov's dog is specifically that our reactions to stimuli are, in at least some cases, mediated through awareness of norms. You could train a dog to believe that there was a fire nearby whenever it heard the sound 'there's a fire': if you repeatedly burned it while saying 'there's a fire', the dog would probably come to react to the sound 'there's a fire' by bolting in terror. But, for all this, would the dog really *understand* the expression 'there's a fire'? Would 'there's a fire' be an expression for the dog the way it is for us?

It seems not. Understanding a phenomenon — e.g. a tokening of a particular expression — consists in deriving its meaning from knowledge of rules or general truths: that is what understanding *is*. Understanding an expression — e.g. 'there's a rock' — is knowing the rules that enable you to derive its meaning. This is obviously true of complex expressions. (For understanding a complex expression involves knowing how its meaning relates to that of its parts; and this involves a grasp of derivation-rules. If you are reacting to 'thereisafire', and not to 'there is a fire', then no matter what your reaction is, you do not understand the sound 'there is a fire'.) But the same is true of simple expressions like 'fire'. If you simply *react* to the sound 'fire', then this sound does not really represent an *expression* for you, even if your reaction is superficially appropriate.[32]

VII. THE MISSTEP IN THE RULE-FOLLOWING ARGUMENT

Why isn't the rule following argument sound? What is the misstep in it? There are two. First, grasping a rule or concept involves grasping how to apply it; one doesn't need to grasp a second rule or concept to know how to apply the first one. So no regress is generated. Knowledge of a rule contains — to some extent *is* — knowledge of how to apply that rule. Anyone who

knows what a stop-sign means knows that, by the time his vehicle is a certain distance from one, he has to have stopped driving. Whenever such a person sees a stop sign, he doesn't need to grasp some second-order rule to know how to apply his knowledge that he should stop; his knowing what the sign means *contains* — in fact, *is* — his knowledge that he should stop. It would be absurd to speak of someone who knew what stop signs meant but didn't know, that when he saw a particular stop sign, he should stop his car. Wittgenstein's regress-of-interpretations argument is generated by assuming that such a person could exist. But such a person couldn't exist; so there is no regress.

Second: *Expressions* are ambiguous. But *meanings* are not ambiguous. (Ambiguity is a property of symbols, not of things symbolized.) One can misunderstand the *sentence* 'grass is green'. But one cannot misunderstand the proposition expressed by this sentence. One either grasps this proposition or one doesn't. The idea that rules can always be misapplied presupposes that rules are always given to one through some kind of language. But must they be? Why? Every person, at some point in his life, has had a thought that he couldn't put into words. How would this be possible if one couldn't grasp propositions *directly*, i.e. not through the medium of language? Of course, if it is possible to directly grasp meanings, then why not rules? (Rules are just a certain kind of meaning.) If one *directly* grasps rules, then these rules are not ambiguous for him; he cannot misapply them. For rules cannot be ambiguous, only the verbal expressions thereof. *A fortiori* we don't have to suppose that public practice is what disambiguates rules — they are not in need of disambiguation.

In light of this point, we can, I believe, prove false a point made by Searle and several others [Searle (1992)]. According to Searle, Chomsky's view that language learning is conditioned by a pre-linguistic grasp of rules is false.

Why? Because, says Searle, it implies a vicious regress. Rules are only given to us in language. So to grasp some set of rules R_1, I must already speak some language L_1. If Chomsky is right, then to speak L_1 I must grasp some set of rules R_2; and to grasp R_2, I must already know some language L_2 (or some 'homunculus' within my mind must know L_2); and so on *ad infinitum.*

Chomsky's view doesn't imply such a regress. For there is no reason, apart from a naturalistic prejudice, to suppose that rules need be given to one through language. So one could grasp R_1 without having to know L_2. 'But doesn't the rule-following argument show that rules need to be given through some language?' No we've already shown the rule-following argument to have no validity — to falsely presuppose that there is a distinction between grasping a rule and knowing how to apply it. So, as far as I can tell, there is no legitimate to reject the idea that we internally represent propositions and, more generally, norms.

VIII. THE FLAW IN BRANDOM'S ARGUMENT

According to both Brandom and Wittgenstein, norms are creatures of *society*. Only after society creates them, do individuals internalize them and come to governed by them. For clarity's sake, let us briefly revisit Brandom's argument.

Let us suppose, along with Brandom, that the mental activity of pre-social creatures is in fact *not* governed by awareness of norms. Then how does a group of such creatures generate (or discover) norms? As Brandom himself realizes, it would not be enough for a group of such creatures to act in regular ways — e.g. to act in unison. The collective behaviour of group of molecules or robots might embody beautiful regularities; but this behaviour wouldn't embody any consciousness of norms. So how would these creature get from having no awareness of norms to being norm-governed? Brandom's answer is this: norms may be *implicit* in the collective behaviour of such creatures.

Individuals come to be aware of norms by being integrated, in some way or other, into these collective practices.

Prima facie this answer is circular; it assumes the existence of norms in order to explain their existence. But Brandom thinks he has a way of avoiding circularity. Imagine the following. Every time it says ‘turn right’, I turn left. Whenever I do this, people beat me with sticks. (So far I am not conscious of norms.) Eventually, I start turning right when it says ‘turn right’. When I do this, I am given a hug. Through this regimen of positive and negative reinforcement, I figure out that ‘turn right’ means *turn right*. Norms were ‘implicit’ in people’s collective behaviour. They beat when I turn left and hugged me when I turned right. Their norms became my norms; I became aware of their norms [Brandom (1998), pp. 34-36].

This account is inadequate: depending on how it is interpreted, it is either circular, as it first appears, or it is based on non-sequitur. As we’ve noted, for Pavlov’s dog, the ringing of the bell doesn’t *mean* anything, at least not in the sense in which ‘dinner is served’ has meaning for us; the dog just reacts to the ringing of the bell. For a physical token to constitute a symbol, one must *already* be in possession of concepts: otherwise one is just reacting like Pavlovian dog to tokens of that symbol. So beating someone with a stick, thereby getting him to turn right when he sees signs marked ‘turn right’, won’t get someone to be aware of norms, unless he already has some latent awareness of norms. (Experience can articulate an existing awareness of norms, but it cannot create an awareness of norms where there was none beforehand.) Behavioural conditioning cannot turn a non-normative creature into a normative one; it can simply turn a creature whose behaviour *obviously* is not norm governed into one whose behaviour is *like* that of a norm-governed creature.

If no individuals are aware of norms, then awareness of norms cannot be

implicit in the behaviour of any aggregate of individuals. A group of normless individuals is a normless group. Of course, this group may act *as if* it were aware of norms. But that is irrelevant. Brandom has avoided the fallacy of regularism in his analysis of the individual's grasp of meaning only to commit this same fallacy in his analysis of the group's awareness of meanings.

> Your dismissal of Brandom's position is much too brief. You say that society cannot be conscious of norms unless some individuals are conscious of norms. But wholes can have properties that none of their parts have. So your argument is based on a non sequitur. No individual neuron is conscious, but certain aggregates of neurons (brains) are conscious.[33] Maybe, in an analogous way, society is conscious of norms even though no individual is (initially) conscious of norms.

So if Brandom's analysis is correct, then societies are individuals in their own right, not just metaphorical or as-if individuals. For the sake of argument, let us grant Brandom this. How then does he explain how these super-individuals become conscious of norms? By Brandom's reasoning, they would have to members of societies in their turn (societies of societies) to become conscious of norms. This would lead to an infinite regress; so it couldn't possibly be correct. In any case, it is empirically false. (Societies are not members of societies of societies. Or even if they are, they certainly don't acquire whatever consciousness they might have by belonging to these second-order societies.) Of course, Brandom could just say that societies are innately conscious of norms. But if he said that, he would be inconsistent with himself. If societies *really are individuals* — and surely they are such if they are aware of things that no individual is aware of (what better criterion of *really being an individual* could we ask for?) — then what Brandom says about individuals (in particular, that individuals must be members of societies

to be aware of norms) applies to societies.

Of course, Brandom might say that societies are not really individuals, except in some irrelevant metaphorical sense. But then, as we have seen, he would have to reject the idea that societies are aware of things that no individual is aware of.

Depending on how he is interpreted, Brandom is making one of two claims. (i) A completely non-norm-governed creature can be turned into a norm-governed creature through some kind of behavioural conditioning that is imposed upon him by his brethren. (ii) Brandom says that individuals become aware of norms because norms are already 'implicit' in the social practices of its species.

Is (i) true? No. A creature that has no grasp of concepts can be made to act *as if* it understood the expression 'roll over'. But it won't really understand this expression; it will just react to it. So if this is the correct interpretation of Brandom's argument, then Brandom's argument is no longer circular, but it is based on a *non-sequitur*.

Is (ii) true? No. As we have seen, if by 'implicit', Brandom means 'non-existent', then his analysis fails. (One 'can't bake a normative cake out of non-normative ingredients', as Brandom himself says). On the other hand, if by 'implicit' Brandom means 'existent in some way or other', then his analysis is circular. Given a group of creatures who are oblivious to norms, be they people or molecules or billiard balls, the behaviour of these individuals, considered individually or collectively, may be *like* that of norm governed creatures; but it won't really be that of norm governed creatures. In order for the group to be aware of norms, at least some of the individuals composing this group must be aware of norms. (Anyway, what does it mean to say that a group is aware of norms except that some of its members are aware of norms?) So, on this interpretation, Brandom is 'explaining' how

individuals can be aware of norms only by assuming, at least implicitly, that some individuals are aware of norms.

In a word, given a group of creatures who are not, individually, aware of norms, there is no way that any of them will ever become aware of norms, no matter how they interact with each other. Awareness of norms is either there *ab initio* or it isn't there at all.

IX. ANOTHER CRITICISM OF BRANDOM

As Brandom himself points out, according to Locke, Kant, and other classical philosophers, speaking is a kind of externalized thinking. Brandom, Wittgenstein, and Sellars say that thinking is a kind of internalized speaking. We have seen reason to believe that Brandom's view is false. Given a creature who is completely oblivious to norms, concepts, and rules, there is no way to get it to become aware of these things: the best that can be done is to instil in it, through behavioural conditioning, a lot of dispositions to act *as if* it were aware of these things. The buck cannot be passed onto society. A society of normless individuals will itself be normless; it will be no different, in the relevant respects, from a bunch of robots. Saying that norms can be 'implicit' in a society of normless creatures is either false or question-begging.

On the other hand, if *per impossibile* a group of normless individuals were, collectively, aware of some kind of norms, then it would have to explained how *this* individual (this super-individual) were aware of norms; and Brandom doesn't explain this.

I have said many times that, unless a creature is *already* aware of rules, a sound cannot constitute an *expression* for that creature. For it is consciousness of rules that makes a mere sound into an *expression* for somebody. But why is this? Suppose that you are not aware of any rules, but that some sound (e.g. 'fire') *makes* you think about fire — as Pavlov's bell

made the dog expect dinner. To use a slightly modified version of one of Wittgenstein's own examples: by ringing a bell or uttering a sound, you could cause a dog to expect dinner *now*; but you couldn't cause it to expect dinner *tomorrow*, let alone three days from now or three days and two hours from now. Ordinary stimuli (by this I mean stimuli that are not tokens of expressions) have no internal structure or articulations. By contrast, sentences are *articulate*, as Wittgenstein himself said early in his career; they have internal structure; they are built up in stages. Sentences are not homogeneous blocks; their parts stand in significant relations to one another; the meaning of the whole is a function of the meanings of its parts. Through a kind of Pavlovian training, you could train a dog to react differently to different stimuli. You could train to expect dinner when it heard the sound 'dinner time', to fear fire when it heard the sound 'fire', to fear choking on smoke when it heard 'smoke', to attack when it heard 'attack'; and so on. But you couldn't compound these Pavlovian expressions into complex parts. If the dog heard 'smoke, fire, attack', it wouldn't be reacting to the *compound* sentence 'there's smoke, and there's fire, and I want you to attack'; it would just be reacting to 'smoke', and then to 'fire', and then to 'attack'; it wouldn't be reacting to one complex expression, but to three separate stimuli. For the dog to hear the compound sentence 'there's a fire, there's smoke, and now I want you to attack', the dog would have to be conscious of the *concepts* of fire, smoke, and attacking. Let me elucidate this.

There is no way to reduce understanding a sentence like 'if there's smoke, there's fire' to having a lot of precognitive reactions. To understand such a sentence, one must grasp, not a particular fire or cloud of smoke, but the *concepts* of fire and smoke. The sentence 'if there's smoke, there's fire' is general; understanding it involves *de dicto*, and not merely *de re*, awareness. The dog can expect dinner *now*, but not *tomorrow*,[34] because expecting

dinner tomorrow involves having a relation to the *concept* of dinner. (It is not a *particular* dinner you expect tomorrow. Expecting *this* dinner is to be analysed relationally, i.e. in terms of your consciousness of some particular dinner; expecting dinner three days from now is to be analysed notionally, i.e. in terms of your consciousness of the *concept* of dinner.) Sentences can have an internal structure precisely because their component parts correspond to *concepts* and not to individuals.[35] Individuals cannot be compounded, only concepts of individuals. 'If there's smoke, there's fire', doesn't express a relation between some actual plume of smoke and some actual fire; it expresses a relation between the *concepts* of smoke and fire. It is not fires and plumes of smoke that are being compounded here, but concepts or meanings. But the dog that hears 'dinner! smoke! fire!' wants *this* dinner, fears choking on *this* smoke, fears being burned by *this* fire. So long as one is always conscious of *this* fire and *this* dinner, one cannot grasp complex sentences about fires and dinners; for *this* fire and *this* dinner cannot be compounded in the way that the *concepts* of fire and dinner can be. (No regimen of conditioning can enable the individual to make the leap from having purely *de re* awareness to having *de dicto* awareness.)

Basically there is no way to understand complex sentence without grasping concepts. To grasp a concept is to grasp a rule. So an understanding of complex sentences presupposes a grasp of rules. This is because to understand a complex sentence you must know how the *meanings* of its constituents are being compounded with one another.

Of course, to understand a complex sentence, one must not only grasp the concepts corresponding to the constituents of this sentence; one must also grasp the rules governing how these concepts are compounded. To understand the sentence 'the bear chased the cat', it is not enough to grasp the concepts *bear*, *chase*, *cat* and to associate these concepts with the appropriate

words. Someone who mistakenly took this sentence to mean *the cat chased the bear* would know all of that; but such a person obviously wouldn't understand the sentence. What such a person lacks is knowledge of the rules governing how sentence-meanings are compounded out of word meanings.

Of course, a Wittgensteinian-Brandomian may say:

> In light of what you've said, I'll admit that a behaviourist analysis cannot account for our understanding of complex sentences like 'if there's smoke, there's fire.' But it can account for our understanding of simple sentence like 'fire' (meaning: there is a fire).

This is obviously untenable. The word 'fire' has the same semantics in 'fire!' as it does in 'where there's fire, there's heat' or in 'kitchen fires should be extinguished with salt, not water'. If the word 'fire' in one word sentences like 'fire!' or 'fire?' affected the listener 'directly', i.e. *not* through the medium of the listener's grasp of some concept; and if the word 'fire' in complex sentences like 'when there's smoke, there's fire' *did* affect the listener through the latter's grasp of some concept; then the word 'fire' would be quite ambiguous. But it isn't.

Also, if the word 'fire', when it occurred alone, didn't operate through the medium of some concept, then it necessarily *couldn't* occur in complex expressions — just as Pavlov's bell couldn't be a constituent of a complex expression. Like Pavlov's bell, it could at most be preceded and followed by other stimuli. It's 'meaning' wouldn't be a part of the meaning, if there were any, of this concatenation of stimuli. (This is simply because psychological and physiological effects, unlike meanings, cannot be compounded, at least not in any sense of 'compounded' that is relevant to semantics.) So in such a case, 'fire' would no more constitute an *expression*, a bit of language, than would Pavlov's bell; it would just be a stimulus. So rules must form the nucleus of any true language. Therefore one cannot grasp a language *until*

one grasps rules. (Wittgenstein's 'language games' thus wouldn't seem to constitute languages at all.)

X. OUR ANSWER TO BRANDOM'S QUESTION

In light of this discussion, we are able to answer a question that Brandom poses at the beginning of *Making it Explicit*. As Brandom points out, a parrot never really says *it's getting hotter* even though it has been trained to say 'it's getting hotter' when, and only when, it is getting hotter. The parrot can produce the *sounds* 'it's getting hotter', but it can't express the proposition *it's getting hotter*. So why can a human being say that it's getting hotter whereas a parrot cannot? This is Brandom's question.

Our answer to Brandom's question is a pretty traditional one. If the temperature is increasing in my room, this may indeed cause me to say 'it's getting hotter'. But this causation will operate *via* my grasp of certain concepts and semantic rules. For me the sentence 'it's getting hotter' has inner structure; it consists of various words, bearing certain markers (e.g. '-ing'), arranged in a certain way; and its meaning is compounded out of multiple different meanings. For the parrot 'it's getting hotter' has no inner structure; it has no inner articulations. (Of course, the physical *token* 'it's getting hotter' has physical articulations. But this is irrelevant.) This inner structure consists primarily in (i) the fact that different constituents of this sentence express different concepts and (ii) the fact that the meaning of this sentence as a whole is some kind of a resultant, and not a mere heaping together, of these concepts.

Our answer to Brandom's question may not be all that different from Brandom's own answer. Actually, it seems to coincide with it:

> Suppose a spectrophotometer is hooked up to a tape-recorder in such a way that it produces a noise of the acoustic type 'that's red' when and only when it is irradiated with the light of the proper frequency. And suppose

> that a fanatical human red-reporter nearly has just the same responsive disposition to produce those noises....What makes the noise the one produces merely a signal on the basis of which someone else might conclude that something red is present, while the very same noise, reliably elicited under just the same circumstances from the other, counts as a noninferential report, expressive of a perceptually acquired belief, with an intentional content the includes the content *red*?...[T]he key element missing from the parrot and measuring instrument — the difference between merely *responsive* classification and *conceptual* classification — is their mastery of the practice of giving and asking for *reasons*[36], in which their responses can play a role as *justifying* beliefs and claims' [Brandom (1998), pp. 88-89].

Basically, I mean *that's red* when I say 'that's red' because I am motivated to produce that sound by my perception of the *appropriateness* of producing them, i.e. because *I know that it is right to produce these sounds*, given my present situation and given the nature of the English language. I grasp some proposition about rules and situations; and because I grasp this proposition I can really *mean* something by the sounds I produce. By contrast, a spectrophotometer doesn't grasp any such proposition; so it doesn't mean anything by the sounds 'that's red'.

This analysis is certainly correct. But it is clear that Brandom cannot give this answer without contradicting himself. I must grasp propositions *before* I can mean anything by words (and, therefore, before words can mean anything to me). But according to Brandom, there is no pre-linguistic intentionality and therefore no pre-linguistic grasp of propositions.

XI. DOES SENTENCE-MEANING PRECEDE WORD-MEANING?

It remains to consider to one last point. I have claimed that sentence-meanings are built up out of word-meanings — or, at least, out of the

meanings of some kind of sub-sentential parts. (Of course, this is a very main-stream, traditional view.) Wittgenstein and Brandom seem to hold that sentence-meaning is primary; that individual terms do not, in and of themselves, have meaning or reference, except derivatively of sentences.[37]

There are couple of reasons to hold this view. First, there is, it seems, no way to understand word meaning except in terms of sentence meaning. A word w can be given meaning only by of giving a rule that assigns truth-conditions to sentences containing w: and this is no less true if w is a 'lexical' item than if it is a non-lexical item.

What does it mean to say that 'Bill Clinton' *refers* to Bill Clinton? What does it mean to say that 'smokes' refers to the property of smoking? It means that the truth of a sentences containing 'Bill Clinton' and 'smokes' depends in a certain way on Bill Clinton and the property of smoking.

Anyone familiar with Tarski-style semantics knows that words are given meaning through rules that assign meanings to the sentences in which those words occur. To say that 'Bill Clinton' refers to Bill Clinton is not to say that 'Bill Clinton' makes me *think* about Bill Clinton. For x to refer to y it is not enough that x make people think of y; guns make people think of death, but a gun doesn't refer to death. To say that 'Bill Clinton' refers to Bill Clinton is, ultimately, to say that sentences of the form '…Billl Clinton…' are true just in case…Bill Clinton….[38] In other words, for 'Bill Clinton' to refer to Bill Clinton is just for *sentences* ascribing a predicate to the expression 'Bill Clinton' to be true just in case Bill Clinton has the property denoted by that predicate. (For example, 'Bill Clinton smokes' is true just in case Bill Clinton has the property denoted by 'smokes'.) And to *understand* the term 'Bill Clinton' is to recognize that sentences of the form '…Bill Clinton…' are thus dependent for their truth on its being the case that…Bill Clinton…Basically,

reference can only be defined in terms of the truth-conditions of *sentences*. So it is reasonable to say that 'Bill Clinton' has no meaning in isolation but only in the context of a sentence.[39]

And to say that 'smokes' refers to the property of smoking is just to say that sentences of the form 'so and so smokes' are true just in case the object denoted by 'so and so' has the property of smoking. So, like 'Bill Clinton', there is no way to assign meaning to 'smokes' except by way of assigning meaning to sentences containing that expression.

Of course, if 'Bill Clinton' and 'smokes' have no meaning outside of a sentence, then 'of' and 'and', and non-lexical items generally, cannot possibly have meaning outside of a sentence. If any kind term had meaning in isolation, it would surely be a referring term. And indeed a term like 'and' cannot be given meaning except by way of assigning meaning to sentences in which it occurs; it must be assigned meaning through some rule like *'S1 and S2' is true just in case S1 is true and S2 is true.*

Second, it seems that one cannot just think about Bill Clinton; one must think about him *as* having this or that property, as wearing a red tie or as being tall. All thought is *predicational*. One cannot think about objects except by thinking about states of affairs involving objects. And this epistemological fact seems to correspond to the linguistic fact we must noted, namely, meaning is assigned to sub-sentential expressions by way of being assigned to sentences.[40]

Given these two points, it would seem that sub-sentential parts *per se* are meaningless and, therefore, that my analysis is misguided.

It may be true that sentences must be meaningful before individual words can be. For the sake of argument, let us grant Brandom this point. But once a term like 'Bill Clinton' occurs in a sentence, it seems to acquire a meaning that is independent of *that particular sentence*, even though it may still have

to occur in *some* sentence to be meaningful. So ‘Bill Clinton’ has no meaning outside of *some* sentence; but it is meaningful outside of any *particular* sentence. ‘Bill Clinton’ obviously has the same meaning in ‘Bill Clinton is tall’, ‘Bill Clinton is angry’, ‘is Bill Clinton coming?’. So ‘Bill Clinton’ has a kind of semantic priority over *any* sentence, even though it cannot have semantic priority over *all* sentences. How could we understand new sentences — like ‘if Bill Clinton had been a more avid athlete in his youth, he wouldn’t have pursued a career in politics with enough zeal to become President’ — unless the words ‘Bill Clinton’, ‘athlete’, ‘President’, and so on, had some meaning that were independent of *this particular sentence* and indeed of any particular sentence that they might occur in?

XII. WHY THERE COULD BE A PRIVATE LANGUAGE

Wittgenstein and Brandom also claim that there could not be a ‘private’ language. (We’ve already considered their reasons for this.) So far we have not explicitly said whether this claim is true or false. But given what we’ve seen, it is clear that we must reject it. In fact, given what we’ve already seen, we must go so far as to say that all languages are grasped privately and that — although there may not in fact be any private languages — there is no important difference between grasping a public language, like English, and grasping a private language. A public language is really just a private language that many people share.

Wittgenstein’s main argument against the existence of private languages is this. A language is a set of symbols whose use is *rule*-governed. A rule cannot be obeyed privately; for a private rule is no rule. To obey a rule, one must be part of a communal practice. *A fortiori* to obey a *linguistic* rule, and thus to grasp a language, one must be a part of a communal practice. So a language is essentially public.

In fact, the existence of these ‘communal practices’ presupposes that the

people who participate in them *privately* grasp the relevant linguistic rules. We've seen that one must grasp propositions *in order* to learn a language. A language is given by a set of biconditionals — e.g. 'P and Q' is true if and only P is true and Q is true. One cannot grasp these biconditionals *through* a language, since it is only *as a result* of knowing them that one can know a language. So one must have non-linguistic access to propositions — in effect, to meanings. Once this is seen, it becomes apparent that in the end *all* rules are grasped privately.

Obviously one's interaction with others — one's being engaged in a 'communal practice' — can *cause* one to grasp the right biconditionals. But it cannot *constitute* one's grasp of these biconditionals. We've seen that given a group of creatures each of whom is individually incapable of grasping a norm, that group of creatures *as a whole* is incapable of grasping a norm. If a group collectively grasps a norm, that is because its members privately grasp norms. To grasp a language is just to have internalized a set of norms. So languages are, in the end, grasped privately.

Of course, languages are, typically, *created* publicly. Presumably the word 'dog' means dog because of the efforts of many interacting people. But the rule *'dog' means dog* must be grasped privately by each individual English speaker. So, in the end, linguistic rules *are* grasped privately: they can *only* be grasped privately. Once this is granted, there is no obstacle to the idea that there could be a private language. Since one's engagement in a communal practice has a *causal*, but not a *constitutive* role in one's knowing a language, it follows trivially that, in principle, a 'society' of one person could have a language: it follows, in other words, that there could be a private language. In fact, each language quite obviously is grasped privately; so, in effect, each 'public' language is really so many private languages: what we call public languages are really idealizations obtained by factoring out the

differences among so many private idiolects; the grammar of a ‘public’ language is a grammar that captures the intersection of so many private languages.[41]

CONCLUSION

There is no doubt that the acquisition of speech enormously enhances and alters one’s cognitive abilities. Some (Wittgenstein, Sellars, Brandom) seem to have concluded from this that the ability to speak a language *constitutes* one’s ability to think.[42] But we have seen that this cannot be so. This leaves us with the question: why does linguistic competence so dramatically increase a creature’s ability to think? Obviously we can’t answer this question here. But my guess is this: speech allows *existing* ratiocinative abilities to be transferred to new domains. Cognitive abilities that are, initially, in the service of quite specific mental operations are permitted, by language, to operate on new contents.

This of course demands elaboration.[JK7] First of all, it is pretty clear that *perception* is the original, and the most basic, form of awareness: more precisely, it is the original and most basic form of *personal* awareness. I have no doubt that there are wholly non-perceptual *sub*-personal processes — e.g. those posited by Chomsky and Marr. But, again, I think that at the level of *personal* mental activity, sense-perception is the prototype for all awareness. And I think that — although, even at the personal level, we obviously have non-perceptual modes of awareness (for example, we have mental images and we also have abstract thoughts) — these are derivatives or sublimations of perception. Now if a creature knows a language, he is able to give his very own thoughts a perceptible form. What we cannot perceive — or cannot access by way of some derivative or sublimation of sense-perception — is no object for us (at the persona level of cognitive activity). So until our thoughts

are ‘clothed’ in a perceptible form, they are no objects for us.[43] So if a person learns a language, his thoughts become *objects* for him; he grasps them, and does not merely have them. I think, in fact, that before a creature knows language, most actually thinking occurs at a sub-personal level. The end-products of that thinking make their way into consciousness; but the think*ing* itself is sub-personal.[44] Once thoughts are given a linguistic dress, the think*ing* can occur at a personal level — for the thoughts have become perceptible and thus accessible to *personal* awareness — and can thus interact with the contents of the personal stratum of mental activity. So the *subject* becomes rational: previously only the *sub*-subject was rational, where by the ‘sub-subject’ I mean, of course, the sub-personal stratum of mental activity. To sum up, the possession of language does not *give* us rationality; it merely increases the scope of our rationality; or, rather, it extends our rationality from the sub-personal to the personal level.

Appendix 1: Why the idea of Sub-personal rule-following doesn’t involve a homunculus fallacy

As we noted, Searle (1992) condemns as incoherent the idea that there is sub-personal rule-following: rules cannot *just* be followed – *someone* must follow them. So if, as Chomsky and Marr maintain, our conscious mental productions result from rules being followed at some sub-personal level, then somebody within us -- some ‘homunculus’ -- must be following them. But if we say that *his* mental productions result from rule following, then we incur a vicious regress. On the other hand, if we *deny* that his mental productions thus result, we are conceding the gratuitousness of the whole idea that rule-following underlies our mental lives.

Searle's argument is compelling, but fallacious; for there can, I submit, be rule-following without a rule-follower. It is clear that, where representational mental phenomena are concerned, the causal order to some tracks the *normative* order. If you believe that Bill has exactly two apples, and you also believe that two is less than three, then you will probably form the belief that Bill has less than three apples. Obviously the first two beliefs not only *cause* but also *justify* the resulting belief. (This point generalizes practically without limit to all cases of 'intentional' causation.) And there can be no doubt that the transition from the first pair of beliefs to the third embodied deference to some logical norm and thus constituted a case of rule-following. But does this mean that some agency *behind* the mental contents was needed to *make* that transition embody awareness of that norm? No. It would seem to be constitutive of representational mental entities that their representational contents are implicated in, and condition, their causal liaisons. If you are thirsty and see water, this will (ceteris paribus) cause you to drink the water, and it will also (logically, though not perhaps ethically) justify your drinking the water. These causal properties of your state of thirst and your perception are inseparable from – are constitutive of – the mental states themselves: given a mental state x that had *no* tendency, when coupled with a perception of water, to result in an intention to drink that water, x would *ipso facto* fail to be a feeling of thirst. So we needn't posit a being *behind* the mental states who *makes* their causal paths track some normative order. No – it is of the essence of representational mental states that their causal properties track nor mative properties and that they belong to sequences of states that embody sensitivity to norms or logical rules, and therefore qualify as cases of rule-following. To sum up, given what representational mental contents are, they cannot but be instalments in cases of rule-following: nothing *behind* them is needed to make this happen. I also

think that, ultimately, nothing behind them *could* make this happen. Such a notion requires a subject who is intimately enough related to mental contents to *have* them, but sufficiently divorced from them that he has some kind of causal powers over them (x can have causal powers over y only if x and y are distinct). And there is, I think, no way to make this idea coherent.

Appendix 2: Why consciousness is not the essence of mentality

Searle (1992) argues that 'non-conscious mental phenomenon' is as much a contradiction in terms as 'four-sided triangle.' Searle's view is not a new one, and was often directed at Freud. Freud produced *empirical* counter-arguments; I'd like to present a purely logical counter-argument.

First of all, the term 'conscious' is ambiguous. To say that x is 'conscious' could mean *either* (i) that one is conscious *of* x *or* (ii) it could mean that x is a *content* of consciousness. To say 'his hatred of his wife is concious' means that he is conscious of this hatred. To say 'pains are necessarily conscious' is to say that pains are necessarily *contents* of consciousness. Obviously when people say that anything mental must be 'conscious', they cannot have sense (i) in mind. For it is obviously viciously regressive, as well as circular, to say that we are conscious *of* everything mental: that would mean that, for any mental entity E, there was some consciousness of E – call it C(E) – and some consiousness of C(E) – call it C(C(E)) – and so on *ad infinitum.* Also, there are obviously things we are conscious of that are not mental (stars, chairs). But people who say that everything mental is conscious cannot have sense (ii) in mind either. For, as everyone will admit, a belief is not a *content* of consciouenss; my belief that 2 is prime is surely not a *content* of consciousness, though I am obviously conscious *of* it. So the thesis that everything mental is 'conscious' cannot

even be delineated without its incoherence being apparent.

References

BRANDOM, R. (1998), *Making it Explicit, Cambridge, Mass.*,Harvard University Press.

CHOMSKY, N. (1990), 'Language and the Problems of Knowledge' in *The Philosophy of Language,* Martinich, A. P. (ed.), New York and Oxford, Oxford University Press, pp.509-527

DENNETT, D. (1978), 'A Cure for the Common Code' in *Brainstorms,* Cambridge, Mass., The M.I.T. Press, pp. 90-108.

DUMMETT, M. (1978), 'Can Analytic Philosophy and Ought it to Be?' in *Truth and Other Enigmas,* Harvard University Press, pp. 437-458

FODOR, JERRY (1975), *The Language of Thought,* Thomas Y. Crowell.

HACKER, G. and BAKER P. M. S. (1984), *Language, Sense, and Nonsense,* Oxford, Blackwell.

MACDOWELL, J. (1994), *Mind and World,* Cambridge, Mass., Harvard University Press.

— (1998), 'Meaning, Communication, and Knowledge' in *Meaning, Knowledge, and Reality,* Cambridge, Mass, Harvard University Press, pp. 29-50.

SEARLE, J. (1992), *The Rediscovery of the Mind,* Cambridge, Mass., The M.I.T. Press.

— (1999), *Mind, Language and Society: Philosophy in the Real World,* London: Weidenfeld & Nicolson.

WITTGENSTEIN, L. (1976), *The Blue Book and the Brown Book,* New York: Harper Torchbooks.

— (1999), *The Philosophical Investigations*, Translated by G.E.M. Anscome, Englewood-Cliffs, Prentice Hall 3rd Edition.[JK8]

—

Emotions and the Content of Moral Statements

Abstract: According to a "realist" analysis of moral statements, "letting children starve" is made true by some moral *fact*, just as "Paris is in France" is made true by some geographical fact and "2+2=4" is made true by some mathematical fact. According to an emotivist analysis, "letting children starve" is tantamount to saying "Boo to letting children starve": it *expresses* one's feelings about letting children starve. It is typically held that a realist analysis of moral statements is incompatible with an emotivist analysis of them. This is the case, I argue, only if emotions are *not* judgments or beliefs of some kind. I also argue that, in fact, emotions are plausibly analyzed as beliefs of a special kind, and that any alternative analysis of them fails to do justice to some of their obvious features (e.g. that we seek *reasons*, and not merely *causes*, for them). This analysis of emotions is also, I argue, the one to which every human subscribes, in his practice if not in his conscious opinions. So emotivism is, I argue, a form of moral realism.

Some sentences are obviously true or false. The sentence "Paris is in Europe" is true, and the sentence "2+2=3" is false. Sentences expressing *moral* sentiments – sentences like "letting children starve is evil" and "you shouldn't encourage your children to smoke" -- are widely thought to be in a different category. According to a theory that has many adherents, such sentences merely express *emotion*. To say "murder is wrong" is essentially to say: "Boo murder!" It is to express a negative feeling about murder. The same is held to be true *mutatis mutandis* of all sentences expressing moral sentiments. To say "you shouldn't encourage your children to smoke" is really just to say "boo encouraging your children to smoke!" Let us refer to this position as *emotivism.*[45]

Of course, many hold that moral sentences are true or false, just like arithmetical or geographical sentences. “Letting children starve is evil” is true and “letting children starve is good” is false, the same being true of all other moral sentences. (By a “moral sentence” I mean one expressing a moral evaluation or sentiment.) Let us refer to this position as “value realism” or, simply, “realism.”

It is typically thought that emotivism and realism are incompatible positions. I think that this is a mistake. I think that they are compatible and possibly even coincident doctrines. It is the purpose of this essay to make a case for this claim.

A disclaimer is in order. I will be able to make, at most, a *conditional* case for the claim just mentioned. *If* a certain reasonable, but highly debatable, doctrine concerning emotions is correct, *then* emotivism is a form of value realism. I will not be able to make an airtight case for that doctrine, but only to provide *some* reason, however minimal, for thinking it credible. But that, by itself, will be enough to warrant our rethinking the relation between emotivism and realism.

First of all, indicative sentences have two functions. They *bear* propositions and they also *express* (or can be *used* to express) beliefs. The sentence “snow is white” bears the proposition *snow is white.* And when someone says “snow is white”, that person is *expressing* his or her belief that snow is white. When I say “snow is white”, I am not *saying* anything about my own mental contents. But I am *expressing* a belief, and thus *showing* you something about my mind.

Sentences are capable of truth and falsity precisely *because* uses of them express beliefs. The sentence “snow is white” is true precisely because uses of it express a true belief – the belief that snow is white. The sentence “Paris is in Russia” is false precisely because uses of it express a false belief. (If

uses of “Paris is in Russia” expressed the belief that 2 is an even number, then “Paris is in Russia” would be true.) So sentences are true or false precisely *because* they express true or false beliefs. [46]

There is more to say in this connection. Many things we do express beliefs. But not all such things are cases of our producing *sentences.* If I put my coat on a table in front of me, I am expressing a belief I have – the belief, roughly, that a physical object in my vicinity is capable of supporting my coat. But my putting my coat on the table is not itself true or false; it is thus not in the same category as my saying “snow is white” or “grass is not green.” The reason that my performing that act doesn’t constitute a case of producing a bit of *language* is that it is not a conventionalized way of expressing a belief.

This no doubt demands clarification. Once again consider the proposition:

(P) *There is an object in my vicinity capable of supporting my coat.*

In English we express (P) with the words: “There is an object in my vicinity capable of supporting my coat.” But we don’t have to use those words to express that proposition. We could invent a code (not a very convenient one, but a code no less) in which the canonical way of affirming P was to put one’s coat on a nearby object. (If one isn’t wearing a coat, then one can’t affirm P in that code – just as a speaker of English who can’t write or say an “s” can’t express P in English.) We could, in principle, make the *conventionalized* way of expressing P be the act of putting one’s coat on a nearby object. If we did that, then one would, in effect, be producing a *sentence* (or, more exactly, a sentence-token) which bore P.

A lot of things we do express beliefs. Not all such things are cases of our producing sentences. When such acts are *conventionalized* ways of

expressing beliefs, they are cases of producing sentences.

The verbiage “snow is white” is a *conventionalized* way of expressing a certain belief. That is why, in producing that sound, one is *saying* something and is thus doing something that can be characterized as true or false. So an indicative sentences is, at bottom, a conventionalized way of expressing a belief.

The emotivist says that moral sentences express emotions. This seems like a reasonable view. It is hard to see how someone could *sincerely* say “it is wrong not to feed starving children” without having some kind of emotion – a negative one -- in connection with the fact that there are starving children. [47]

This doesn’t mean that whenever one makes a moral statement one is in the throes of emotion. If I say “letting children starve (when there is plenty of food to give them) is evil”, I don’t necessarily have to be in a fit of rage or of outrage. But emotions are not always *occurrent.* I can hate somebody for forty years and not always be in a fit of anger. The emotion shouldn’t be confused with the occurrent manifestations of that emotion. Now if somebody felt nothing – had *no* emotion in connection with the phenomenon of children needlessly starving – then it is hard to see how that person could *sincerely* say “letting children starve is evil”.

I will operate on the assumption that emotions *do* express emotions (though not necessarily occurrent ones). I admit that this is an arguable point. But it is, I think, reasonable enough to warrant our investigating the consequences of it. And one of those consequences, I submit, is that Emotivism is a form of moral realism.

So let us accept, if only provisionally, that to *sincerely* say “letting children starve is evil”, one must have some kind of emotion in connection with letting children start.

Now it is hard to deny that moral sentences are *conventionalized* ways of expressing emotion. The sentence "we are morally obligated to feed starving children" is obviously a *conventionalized* way of expressing a certain sentiment. It is, in this respect, very different from a natural or non-conventional way of expressing sentiment. If one were to gasp in horror at the spectacle of some starving children, that would, perhaps, express one's sentiment that starving children ought to be fed. But that performance would not a natural or non-conventional way of expressing that sentiment, and would thus not be in the same category as one's *saying* "starving children ought to be fed."

To sum up, it is very hard to deny that moral sentences are conventionalized ways of expressing emotions. In any case, for the sake of argument, let us assume that this is correct, and see where it takes us.

Given this supposition, if emotions *are* true or false beliefs of a certain kind, then sentences expressing emotions, and thus moral sentences, are themselves true or false. If "we are morally obligated to feed starving children" is a conventionalized way of expressing an emotion, and if that emotion is a true or false belief, then that sentence is true or false.

To sum up: If emotions are beliefs, then moral sentences – sentences like "letting children starve is wrong" – are true or false, just like sentences expressing geographical or arithmetical propositions. And in that case emotivism actually becomes a kind of moral realism.

The question, then, is: *Are* emotions beliefs? This is a large question, and of course it cannot be disposed of in a few pages. But it is, I think, possible to outline a prima facie case for this position.[48]

If you are angry that Bob stole your bicycle, that means not merely that you *believe* that Bob stole your bicycle, but also that you have some kind of attitude *towards* that fact. Clearly if you didn't have an *attitude* towards that

fact, you wouldn't have any *emotion* in connection with it. What is it to have an attitude towards a fact? Well, it is *not* merely to have a feeling that is *caused* by that fact. Anything can cause anything. Suppose that, unbeknownst to myself, I inhale small amounts of some gas that subtly alters my brain-chemistry, causing me to feel elation. My inhaling the gas *causes* me to feel elated, but I am not elated *about* inhaling the gas: I did not even know that I did so. In general, to have an emotion about x is not the same thing as to have a feeling that is *caused* by x.

To have an emotion about something is to have an attitude towards that thing. To have an attitude towards a thing is to feel a certain way about that thing. If a feeling is to be *about* something, it must *represent* that thing as being a certain way. Barry Stroud says that all perception is "predicational"[49]; one cannot just see a thing; one must see it *as* having these or those properties – as having a certain shape or being a certain distance from one. This point generalizes to all mental representations. A thing cannot *just* be represented; it must be represented *as* having this or that property, *as* standing in this or that relation.[50] In particular, a feeling cannot *just* represent a thing; it must represent it *as* having this or that property. To have a feeling that represents a thing is to have a feeling about a thing. To have an emotion about a thing is to have a feeling about that thing. So given what we said a few sentences a go, it follows that in having an emotion, one represents a thing *as* being a certain way – as having some property or standing in some relation. So to have an emotion is, in effect, to see a thing as having a certain property or standing in a certain relation: it is thus to have a belief to the effect that that thing has some property or stands in some relation.

Of course, many philosophers will resist any analysis that turns emotions

into beliefs.[51] We feel that, in some way, emotions are the *opposite* of beliefs. A belief is what you *don't* have when you have an emotion. In any case, a *good* belief – an objective belief – is what you don't have when you are emotional. A belief is a window onto reality; an emotion is like tar smeared on that window.

It is true that emotions do cloud our judgment. But that doesn't necessarily mean that they are non-beliefs. It could be that emotions embody the functioning of a more retrograde and primitive dimension of mental functioning than do, say, mathematical or geographical beliefs. So emotions might render opaque our epistemic window onto reality *not* because they are not themselves beliefs -- *not* because they are not themselves such windows – but because they are beliefs that result from a less high-grade form of cognition. (Our world view is as much distorted by false or simplistic beliefs as it is by non-beliefs.)

Let us consider a point made by Donald Davidson. Every rational human being knows that wrenching and twisting your body *after* rolling a balling bowl will not make it more likely to land a strike. But just about every human being engages in such wrenching and twisting. What is going on here is a conflict within the bowler's belief-system. Speaking roughly, the bowler's reptile-brain believes that wrenching and twisting *will* make the ball land a strike. The bowler's fore-brain knows otherwise. The belief formed by the reptile-brain clouds the bowler's judgment. But, notice, what is doing the clouding is a *belief*.

Basically, *beliefs* can cloud beliefs. Irrational beliefs can retard the generation or execution of more rational beliefs. So given only that emotions fog up our belief-window onto reality, it doesn't follow that they are not themselves beliefs. I will return to this point later.

Right now I'd like to give a more positive reason to think that emotions

are beliefs. The argument consists of two steps. The first step is uncontroversial.

Emotions obviously *incorporate* beliefs. If you are angry that Bob didn't do the dishes, that means you *believe* that he didn't do the dishes. If you are afraid that Mary will kill you, that means you believe there to be a non-negligible probability that she may kill you. (It doesn't mean, obviously, that you *know* she will kill you. But it *does* mean that you have *some* beliefs concerning the *probability* of such an occurrence.) If you hope that Mary will *not* kill you, you *believe* there to be a possibility of her not killing you. (If you *know* that she will kill you, you can't really *hope* that she won't kill you. If you *know* that P, you cannot hope that not-P. I *know* that my best friend just died. Therefore I cannot *hope* that he hasn't died.) Uncontroversially emotions all *involve* beliefs. This much must be granted.

This brings us to the second step of the argument. Given someone who says that emotions are *not* beliefs, that person must say that emotions merely *involve* them – that beliefs are a proper part of emotions, but are not the whole of them. So such a person will say that an emotion is a belief *plus* something else. What is that something else? Well surely it is a *feeling*. If you are angry that Bob didn't do the dishes, then you believe that he didn't do the dishes. *But in addition to that* you have some kind of feeling. (If you knew that Bob didn't do the dishes, but had *no* feelings at all – if you didn't care – then you wouldn't have any emotion about his not doing the dishes.) And the emotion is the sum of the belief plus the feeling.

But what kind of feeling must it be? It must be a feeling *about* Bob's not doing the dishes. If you know that Bob didn't do the dishes, and you have some feeling *about* something else – say, about the fact that you didn't get tenure – then you don't have an emotion about *Bob's not doing the dishes.* So the feeling in question must be *about* Bob's not doing the dishes. But one

thing cannot *just* be 'about' another. A book cannot *just* be about Spain. It must say something about Spain. A dream cannot just be about your mother. It must have her riding on a broomstick or baking cookies. Remember what Stroud says: things cannot just be represented; they must be represented *as* having these or those properties. So if that feeling is *about* Bob's not doing the dishes, it cannot *just* be about it; it must, as it were, say something about. Your feeling is *to some effect.* Presumably, it is to the effect that it was *wrong* of him not to the dishes or that he failed you by not doing them.

In any case, the feeling must be *about* Bob's not doing the dishes. So, in effect, it must represent Bob's not doing the dishes. But it cannot *just* represent Bob's not doing the dishes; it must represent his not doing the dishes *as* having these or those properties. So the feeling must be a feeling *to the effect* that Bob's not doing the dishes has these or those properties – say property phi.. So the feeling says: Bob's not doing the dishes has phi. But in that case, it is very hard to see how that feeling is anything other than a belief. The feeling is one to the effect that Bob's not doing the dishes has phi. So in having that feeling, you see Bob's not doing the dishes as having property phi. So you believe that Bob's not doing the dishes has phi.

Let us sum up. Emotions obviously involve beliefs; if you are angry that Bob didn't do the dishes, you obviously believe that he didn't. But that belief by itself is not an emotion. Something – a *feeling* – must be added to it. But that feeling must be *about* the right thing – it must be about Bob's not doing the dishes. (Otherwise you will an emotion about something else, or perhaps you won't have an emotion at all.) But if the emotion is about Bob's not doing the dishes, then it represents that fact in a certain way – as having these or those properties. So in having that feeling, you *see* Bob's not doing the dishes as having some property, and thus *believe* his not doing the dishes to have that property. (Later on, we will discuss what kind of property that must

be.) Given any emotion, an exactly analogous argument shows that it is a belief.

Attempts to turn the emotion into something other than a belief are failure. Such attempts turn my anger that Bob didn't do the dishes into the sum of two things: (i) a belief that he didn't do the dishes And (ii) a feeling of some kind. No one can deny that my anger incorporates such a belief. So if we are to turn emotions into non-beliefs, we must see to it that (ii) – the *feeling* part of my anger that Bob didn't do the dishes -- is something other than a belief. But if the feeling part isn't *about* Bob's not doing the dishes – if it doesn't represent his not doing the dishes – then the sum of (i) and (ii) won't be an *emotion* about Bob's not doing the dishes. On the other hand, if (ii) *is* about Bob's not doing the dishes, then we end up with a belief. So, we've seen at least one good theoretical reason to hold that emotions are beliefs.

This analysis corresponds to our pre-theoretical intuitions. If I am not mistaken, we pre-theoretically believe that our emotions are insights of sorts. Everyone wants to get rid of an unpleasant feeling, such as a headache or a twinge in one's elbow. But people cling to their emotions, even the most unpleasant ones. A person who is violently angry towards somebody doesn't want to take a pill to get rid of his emotion, even though that anger may be extremely distressing. If I am angry that my child was kidnapped, what I want is *not* to directly get rid of the emotion, but to change the world so that the emotion is no longer justified. If somebody kidnaps my child, the last thing I would choose to do is to take a pill, or undergo an operation, that would get rid of the emotion. What I want is to change the world so that the emotion is no longer appropriate: I want my child back; I want the kidnapper brought to justice. This suggests that, in having the emotion, I am seeing the world as being deficient in some way – as being such that it ought to be changed. And this, in turn, means that in having the emotion, I believe the

world to have a certain property – as being deficient and in need of change.

Once again, suppose that my child has been kidnapped. Naturally, I am enraged and distressed beyond all belief. But because of recent developments in psycho-active medications, I have the option of taking a pill that will actually make me *happy* that my child was kidnapped. Taking the pill will involve no effort at all. At the same time, I also know that, if I make some effort and go to some expense – if, say, I hire a detective and, perhaps, do some detective work myself -- I stand a good chance of getting my child back. What will I do? Any human being would choose to make the effort – to endure some distress and expense – and get his child back rather than to take the pill.

If you take a pill to get rid of a head-ache, you are not *reducing* yourself in any way. If you take a pill to get rid of your desire to find your lost child, you are reducing yourself; you are narrowing the scope of your consciousness. In any case, this is how we feel about our emotions. We part with them – with our hatreds, our loves, our regrets -- only grudgingly, only when, in the light of contrary evidence, we feel them to be unreasonable.

This suggests that we do not see emotions as being merely subjective *conditions* like headaches, but as having some meaning. No one will hesitate to take a pill to get rid of a headache. For we do not feel that, in getting rid of the headache, we are blinding ourselves to some kind of truth. But, in our view, the parent of the kidnapped child would be blinding himself to some meaning, and also succumbing to some delusion, if he took a pill to convert his misery into joy. So, we feel, the anger is a genuine *recognition* of some kind, and the joy that the parent *would* feel, if he took the pill, would be a *delusion* of some kind. It thus seems to us that, in having the anger, the parent is seeing things the right way – is having the right belief – and that, in having the joy induced by the pill, the parent is seeing things the wrong way – is

having the wrong belief. So we have very strong pre-theoretical intuitions to the effect that emotions are *recognitions* of meanings or truths of some kind.

Of course, it is true that people take pills to change their feelings. But it seems to me that we do that as a last resort: what we *really* want is to change the world in the way consistent with the insight that is, we believe, embodied in the emotion. A depressed person's *initial* desire is to change the world in such a way that the emotion is no longer appropriate. If I am unhappy because I did not get tenure, what I *initially* want is to change the world in such a way that the emotion is no longer appropriate: I want some (perceived) deficiency on the part of the world to be corrected. (I want to get tenure, or I want the people who denied me tenure to be punished.) But I may know that that will not happen (the tenure committee simply will not revise its decision, nor will they be brought to justice). So, under that circumstance, I may indeed take Prozac. But I do that only because other, more desirable options are closed: I want to change the *world*; since I cannot do that, I change my biochemistry.

Also, when we *do* decide to take a pill to change our feelings, that is often (though not always[52]) because we believe that the emotion in question – the one we are trying to get rid of – is *irrational*. At first this might seem to bear against the contention that we see our emotions as being beliefs, or insights, of a certain kind. But the opposite is the case, as a little analysis will show.

No one thinks that a headache is *irrational.* Headaches are unpleasant, but not irrational. But emotions can be irrational. In any case, this is what we tend to think. We describe people as prone to irrational emotional highs and lows. I am not saying that we are necessarily *right* to speak this way. I am merely registering the fact that we do speak this way – that human discourse is replete with such talk. And this fact surely suggests that, at some level, we

see emotions as somehow being answerable to canons of rationality and thus as being quite *unlike* headaches, which we do not see as being irrational or rational. We don't say that people are prone to irrational headaches or stomachaches. Of course, something cannot be answerable to canons of rationality – cannot be rational or irrational – unless it is a belief. So, it would seem, at some level, we see our emotions *not* as things merely undergone, like headaches, but as true or false insights, as *beliefs.*

People *defend* their emotions. People justify their anger or hatred.[53] People give *grounds* for their emotions; and this suggests that people take their emotions to be true or false, and thus to be beliefs. People do not justify, and do not give *grounds* for, their headaches.

It must be pointed out that giving *grounds* for something is different from giving the *causes* of it. I can give the *causes* of my headache (having too little caffeine or too little sleep); but I cannot give *grounds* for my headache. But I can give *grounds* for my hatred of so and so or my delight that such and such. If grounds can be meaningfully given for emotions, then they are true or false, right or wrong. And any mental entity that is right or wrong is *ipso facto* a belief.

So it certainly appears that emotions can be *justified*: that we can do more than give their *causes*; we can give *grounds* for them. To give *grounds* for something is to show that it is rational. The only things that can be rational or irrational are beliefs. So if – as we all pre-theoretically believe – grounds *can* be given for emotions, then emotions are beliefs.[54]

We've seen a theoretical reason to hold that emotions are beliefs of some kind, and we've also seen that we pre-theoretically see emotions as being beliefs of some kind. So the idea that emotions are beliefs is not to be dismissed lightly. And, as we've also seen, *if* emotions are beliefs, then moral

sentences are true or false; and emotivism becomes a kind of value-realism.

We've spoken a great deal about why emotions *are* beliefs, and also about how this view has deep roots in our pre-theoretical outlook. But we have not spoken about another aspect of our conception of emotions. Although, arguably, emotions are beliefs, they *appear* to be very different from paradigmatic beliefs, like the belief that snow is white or the belief that 2+2=4. I think that this appearance has led many to see emotions as *not* being beliefs: given how different anger is from, say, an arithmetical belief, we must surely hold that anger is not a belief.

There *is*, I grant, a crucial difference between emotions and, say, arithmetical or geographical beliefs. But this difference is between *kinds* of belief, not between beliefs and non-beliefs. The belief that Paris is in France is purely "factive": it is *not* a belief to the effect that something is good or bad, that something has or lacks value. The same is true of arithmetical beliefs, historical beliefs, and economic beliefs. Arguably, it is true of *anything* that is *uncontroversially* a belief.

But *if* emotions are beliefs – and we've seen some reason to think that they are – they are *valuative* beliefs, and not purely factive ones. Given any purely factive belief, one can, in principle, have any emotion in connection with it. My believing that 2+2=4 doesn't necessarily involve my having any emotion at all. In fact, even my believing that Stalin killed 10 million people doesn't *necessarily* involve my having this or that emotion. It is true, given the psychological make-up of any ordinary human being, that one is *likely* to experience horror or outrage in connection with such a belief. But this is still a contingent fact about certain human beings. It is not *necessarily* the case that one has any particular emotion in connection with that belief. In fact, there are probably are people who, dreadful as it sounds, have little or no emotion in connection with that belief (or with beliefs concerning similarly

horrible truths). So no factive belief is such that, in virtue of having it, one necessarily has any particular emotion. Thus, if emotions are beliefs, they must be *valuative* beliefs. An emotion, then, must be a belief to the effect that something is good or bad.

We have seen some theoretical grounds for this last contention. And, I think, it is very much consistent with our pre-theoretical intuitions. My being angry at so and so for doing such and such seems to involve my believing that it was *wrong* of so and so to do such and such – that I was slighted, not given my due. The same point holds *mutatis mutandis* for all other emotions. My being happy that such and such happened seems to involve my seeing such and such as *right.*

There are some apparent counter-examples to this thesis. I have said that if a person has a positive emotion towards something – if, say, he delights in it – that means that he sees it as good. Now a person may delight in such and such, but feel guilty about delighting in it, and thus think that he is *wrong* to delight it in. For example, a person may delight in some misfortune that happened to someone else, and also think he is *wrong* to have that feeling. And he thinks it is wrong to have that feeling because, presumably, he thinks that it is *bad* that the misfortune happened to that person. So, if my analysis is correct, then a person simultaneously sees a certain thing as good *and* as bad. Offhand, this seems to bear against our analysis. People do not see a given number as being *both* even and odd. Similarly, people don't see acts as being *both* good and bad. Surely people are not *that* irrational

It is, I grant, a simple fact that people often think themselves wrong for having certain emotions. But this doesn't necessarily bear against our analysis. For, it seems to me, what is going on in those cases is that one has conflicting beliefs as to the goodness or badness of the state of affair in question: people *do* have contradictory views – and the error in the argument

given in the last paragraph lies in its denying this obvious truth.

This demands amplification. First of all, it is hard to shake off the conviction that, in having an emotion in connection with something, one is *at some level* having a belief as to the goodness or badness of that thing. I cannot be angry at somebody without *at some level* believing that he has somehow fallen short. I cannot delight in something without *at some level* seeing that thing as good.

It seems that, in some cases, our forebrains (if I may so speak) generate beliefs that are at variance with the beliefs generated by our back-brains. (Here, of course, I am using the term "forebrain" as a label for that cognitive system associated with rationally produced beliefs, and "back-brain" as a label for that cognitive system associated with instinctively produced beliefs.) Our forebrain tells us that contorting our bodies after rolling a bowling ball will have no influence on whether that ball goes into the gutter or not. Our back-brain, it would appear, tells us otherwise. As Davidson pointed out, even the ablest of philosophers will wrench and contort his body after rolling the ball, in an apparent to influence its course.[55] Our forebrain tells us that astrology is nonsense. Our back-brain tells us otherwise: that is why, if the astrologer tells us we have two weeks to live, we shudder in horror (unless we are extraordinarily disaffected people). It could be that, when a person delights in something that he condemns, or is unhappy about something the he approves of, what is going on is a similar conflict between forebrain and back-brain. My back-brain sees it as a good thing that my department head just fell off a cliff. My forebrain sees it as a bad thing. (My forebrain has been enculturated, and has possibly even read Kant and Mill. None of this is true of my back-brain, which thus has much a more egocentric understanding of the concepts of goodness and badness.) So when I simultaneously delight in his falling of a cliff *and* see myself as wrong to experience that delight,

what is going on is *not* that I *don't* believe it good that he fell off a cliff: rather, I simultaneously see it as *bad* that he fell of a cliff – this is the forebrain at work – and also as good that he fell off a cliff – this is the back-brain at work.

(All this talk of "forebrain and back-brain" may sound like pseudo-scientific bunk. But it just helps us state, in an abbreviated way, things that practically everybody countenances. We say things like: "Smith thinking with his heart, not his head: he really shouldn't give all his money to Charlene" Obviously, Smith isn't literally thinking with his heart. But what we mean when we say things like "Smith is thinking with his heart…" are often true. The same applies to my talk of "forebrain" and "back-brain". It is just an analogue of our often veridical, though anatomically under-informed, talk of "head versus heart".)

Once again, these remarks are rough and ready. But they provide a not entirely implausible way of reconciling the fact that we often condemn our emotions – see them as wrong -- with the (as we've seen) very reasonable thesis that emotions are beliefs as to the rightness or wrongness of things.

Conclusion

What makes a sentence true or false is, ultimately, that it expresses a potential belief. According to emotivism, moral sentences express emotion. If emotions *are* beliefs, then moral sentences are true or false; and emotivism is a form of value realism, contrary to what has been maintained for centuries. There is some reason to think that emotions are beliefs. An emotion is necessarily *about* something. Consequently, it must *represent* a thing as being a certain way. Now if an emotion represents a thing as being a certain way, then it is a *belief* to the effect that that thing is that way. Also, people

give *grounds* for their emotions. So people see their emotions as answering to canons of rationality – as being rational or irrational. The only things that can be irrational or irrational are beliefs. So if – as everyone pre-theoretically believes -- emotions are rational or irrational, then emotions are in fact beliefs.

References

Blackburn, Simon

[1] *Spreading the Word: Groundings in the Philosophy of Language*. New York: Oxford University Press, 1984

[2] *Ruling Passions: A Theory of Practical Reasoning* New York: Oxford University Press, 1998

Calhoun and Solomon

[1] *What is an Emotion?* Oxford University Press: 1984 edited by Cheshire Calhoun and Robert Solomon.

Hume, David

[1] *A Treatise of Human Nature* ed. Selby-Bigge; 2nd edition with text revised and variant readings by P.H. Nidditch. Oxford at the Clarendon Press 1978.
[2] Hume, David. *Inquiry into the Principles of Morals.* Oxford University Press, 1998

Why conceivability sometimes entails possibility: a Neo-Cartesian argument for dualism

Abstract: Kripke and Putnam showed that what can coherently be conceived of is not *always* a guide to what is possible. But in *some* cases conceivability is a guide to possibility: whether it is or not depends on the structure of the proposition being considered. I make it clear under what conditions possibility and conceivability coincide, and I then show that that, where the propositions relevant to the mind-body problem are concerned, they *do* coincide. Thus, given that mind-body dualism is *conceivable*, it follows that it is *actual* – even though, as Kripke and Putnam were right to say, conceivability is not *generally* a guide to possibility. Of course, if dualism is correct, then the validity of physicalistically oriented research programs in psychology would have to be reevaluated.

Introduction

It used to be thought that the possibility, if not the actuality, of some kind of mind-body dualism could be established through a *conceivability* argument. Kripke and Putnam[56] showed that conceivability is not nearly as reliable a guide to possibility as used to be thought: I can coherently conceive of x as being water and also as *not* being H_2O: but if x is water, then it is necessarily H_2O. So what can coherently be conceived of is not necessarily possible. Thus, even though I can coherently conceive of mind as *not* consisting of (say) patterns of neuro-chemical activity, it doesn't follow that mind *could* be anything other than such activity – let alone that it *is* something other than such activity. Understandably, this line of thinking has

convinced many that dualism cannot be established by means of conceivability arguments.

In this paper I will try to show that – even though Kripke and Putnam were basically right in what they said about conceivability as a guide to possibility – it is *still* possible to construct a viable conceivability argument for dualism.

The classical conceivability argument

The old argument for dualism was basically this.[57] What is not contrary to the laws of logic is possible. Triangles can be green because *x is a triangle* is logically consistent with *x is green.* Now *x is a pain or a belief that 2+2=4* does not logically entail *x is a brain event.* So it is logically possible that beliefs, pains, etc. should be distinct from brain-events (or any other kind of physical event.).

The next step in the argument is this. Identity holds *necessarily.* If A *can* be distinct from B, then A *must* be distinct from B. Proof: B obviously doesn't have the property that it can be distinct from B. So if A has the property that it can be distinct from B, then A has a property that B does not have and so, by Leibniz's law, A is simply *not* B.[58]

To sum up the argument so far: (i) Identity is necessary. (ii) *x is a pain (or a belief that 2+2=4…)* doesn't entail *x is a brain-event*. (iii) If P is not ruled out by logic (i.e. if not-P is not logically necessary), then P is possible.

(i)-(iii) imply that pains, beliefs, etc. are not brain-events: dualism proved.

These days, of course, the counter-argument is to deny (iii) – it is to deny that logical possibility entails actual ('metaphysical') possibility (single-spaced statements are those of an imaginary objector):

Given only that *x is a pain (or a belief that 2+2=4…)* is logically consistent with *x is not a brain event*, it does *not* follow that pains are necessarily *distinct* from brain events. Why not? Well, *x is water* is logically consistent with *x is not* H_2O, but we know from chemistry that water *is* H_2O. Some necessities are *a posteriori*, and logical possibility therefore proves nothing as to *actual* ('metaphysical') possibility.

The conclusion, I submit, is vastly overdrawn. Surely logical possibility *sometimes* indicates actual possibility. *x is a triangle* is logically consistent with *x is green*, and this *does* mean that there could be green triangles. At the same time, of course, *x is Hesperus* is logically consistent with *x is not Phosphorous*, but nothing that is Hesperus could not be Phosphorous. And *x is water and x is not* H_2O is logically consistent, but this does *not* mean that water could be something other than H_2O. What is the relevant difference among these cases?

The relationship between conceivability and possibility

Now I would like to state, without giving any argumentation, my position regarding the relationship of conceivability to possibility. I will argue for this position in the next section.

Whenever there is a disparity between logical possibility and actual (metaphysical) possibility, or between metaphysical necessity and logical necessity, that is because the possibilities/necessities in question are underwritten by facts about the structure of *object-dependent concepts.* And whenever a necessity is underwritten wholly by facts about the structure of object-*in*dependent concepts, actual necessity/possibility coincides with logical necessity/possibility.

(In this paper, I will use the word ‘concept’ to denote a certain kind of *platonic* entity: the meaning of a predicate, essentially. I will not be using it to denote concepts in the psychological sense. So I will use the word ‘concept’ in the sense it has in the sentence ‘for any x, if x falls under the concept *square*, x falls under the concept *closed figure*.’)

Before we continue, we must make it clear what is meant by the terms ‘object-*de*pendent concept’ and ‘object-*in*dependent’ concept. Consider the proposition.

(*)*Socrates drank hemlock*

Socrates himself is an actual constituent of (*). But what does *this* mean exactly? The idea will become more clear if we consider a proposition that does not have Socrates for a constituent (i.e. that is not ‘de re’ about, or ‘object-involving with respect to’, Socrates):

(**) *there was a philosopher of antiquity who exceeded all others in philosophical ability and any philosopher answering that description drank hemlock.*

(**) is *made true* by the fact that Socrates was the greatest philosophers of antiquity and that he drank hemlock. But (**) does not have Socrates himself as a constituent. That very proposition does not depend for its truth on *Socrates’* having those characteristics. (**) would have been true if Socrates had never existed, and *some other person* had been the greatest philosopher and drank hemlock. So Socrates himself is not a constituent of (**).

By contrast, (*) would *not* be true if *anyone other than Socrates* had the property of being the greatest philosopher of antiquity and of drinking

hemlock. (*) depends for its truth on Socrates -- specifically Socrates -- having those properties. So Socrates himself figures in the truth-conditions of (*). And, in as much as propositions are internally or essentially related to their truth-conditions, Socrates himself can be said to be a constituent of that proposition.

Now it seems reasonable, if not truistic, to say that propositions are built out of concepts and, therefore, that (*) is constructed out of certain *concepts*, these being *Socrates, hemlock,* and so forth. Given this last fact, and given – what we saw a moment ago – that Socrates himself is a constituent of (*), it seems to follow that the concept *Socrates* has Socrates himself for its content, or at least for a part of its content. Obviously Socrates manages to be a constituent of (*) only by way of his involvement in the concept *Socrates.* So the concept *Socrates* has an actual constituent of the spatio-temporal world for its content (or at least a part of its content), this constituent being Socrates. So the concept *Socrates* is *object-involving*, as it is generally put (it is object-involving *with respect to Socrates*).

(If a concept has only a platonic object for its content, it does not count as 'object-involving' – at least not as I will be using this expression -- even though platonic objects are presumably objects of sorts.)

Object-involving concepts can have natural kinds for their contents – their contents needn't always be spatio-temporal *individuals*. Consider the proposition

(***) *water freezes at 32°*

This proposition is object-involving with respect to the natural kind *water.* What does this mean? The best way to see what this means is to contrast it with a proposition that is not object-involving with respect to water:

(****) there is some substance that human beings bathe in that freezes at 32°.

(****) is *made true* by the fact that we bathe in water and that water freezes at 32°. But (****) doesn't depend for its truth on *water*'s having a certain freezing point. If there were some other substance with a freezing point of 32° that we bathed in, then (****) would be true. But (***) is not like this: for (***) to be true, it is necessary that *water* – specifically *water* -- be such that we bathe in it and that it have a certain freezing point. So *water itself* figures in the truth-conditions of (***). And in as much as propositions are internally or essentially related to their truth-conditions, water itself – the natural kind -- can be said to be a constituent of (***).

To echo what we said a moment ago: It seems reasonable, if not truistic, to say that (***) is built out of various *concepts*, these being *water*, *32°*, and so forth. Given this last fact, and given that water itself – the natural kind – is a constituent of (***), it seems to follow that the concept *water* has the natural kind *water* for its content, or at least part of its content. Presumably that natural kind succeeds in being a constituent of (***) only by way of its involvement with the concept *water*. So that concept must have the natural kind *water* for its content. Thus, the concept *water* is *object-involving*; for it has a natural kind for its content.

The concept *triangle* is not object-involving; for its content is some purely platonic object, not an individual, and not some natural kind like water. The same is true of various other concepts: *number, justice, truth, knowledge, implication.* These concepts do not – so my intuition tells me -- have spatio-temporal individuals or kinds for their contents; they are not like the concepts *Socrates* or *light.* (Later on I will discharge that intuition.)

When a proposition is necessarily true, and its truth is underwritten

wholly by facts about the structure of object-*in*dependent concepts, that proposition is a priori. Consider the proposition *squares have four sides.* The truth of this proposition is guaranteed by the structure of object-independent concepts, and that is why it is a priori.

Of course, I am taking it for granted that *four*, *square*, and so on, are object-independent. In a little while, I will defend that assumption. But it is clear, I think, that *four* and *square* are in a different category from *Socrates* and *water*. Facts about the essences of the first two concepts are epistemically transparent, while the corresponding facts[59] about the latter are epistemically opaque: capable of being known only through empirical investigation. If I want to learn what it is for something to be a square, I do mathematics: I engage in a priori reasoning; I don't do empirical work. If I want to learn what is for something to be water, or if I want to learn what properties Socrates has in all possible worlds, I must do empirical work; I must look into, say, Socrates' conditions of origination; I must investigate the chemical structure of water. So what I investigate to learn about the essence of the concept *square* is not something that belongs to the empirical world; it is something that belongs to a purely platonic realm. Otherwise I *would* have to do empirical work to learn what it is for something to be a square: but I don't. But I *do* have to do empirical work to learn what it is for something to be water; and that, surely, is because that concept is object-*de*pendent; it has an actual spatiotemporal entity for its content (or at least part of its content); and spatiotemporal things can be known only through empirical work. So it is at least *prima facie* plausible to suppose that *square* is object-independent – doesn't have for its content any constituent of the spatiotemporal world – while *water* is object-dependent. We will develop this argument shortly.

When a proposition is necessarily true, and its truth is underwritten by the structure of object-*de*pendent concepts, that proposition is a posteriori. The

truth of *Hesperus is lovely* necessitates the truth of *Phosphorous is lovely.* But this necessity is underwritten by object-dependent concepts (*Hesperus, Phosphorous*) is therefore a posteriori; it is not a *logical* necessity, it is not an entailment. The same is true *mutatis mutandis* of the necessary connection between *x is water* and *x is* H_2O.

On the other hand, the concepts of triangularity and of two-sidedness are not object-involving, and that is why the necessary relation between *x is a triangle* and *x has more than two-sides* is epistemically transparent – that is why one can know *a priori* that, if x has only two sides, then x is not a triangle.

Our argument for this position

In the previous section we said the following two things. First, where object-*in*dependent concepts are concerned, logical possibility *does* correspond to actual possibility. But, second, this is *not* the case where object-*de*pendent concepts are concerned.

We've considered specific examples of necessary propositions (*squares have four sides*, *water is* H_2O), and what we found certain bore out the two contentions just stated. But that fact doesn't constitute anything resembling a general *justification* or proof of those propositions. Is there some such general justification?

There is. By definition, object independent concepts do not have any part of the spatio-temporal world for their contents. Now whether some kind of necessary relation holds between two concepts C1 and C2 has to do entirely with the constitutions of those concepts. Why? Necessary relations hold in all possible worlds. So whether a necessary relation holds between two concepts

cannot be contingent on what is the case in this or that world; it must, therefore, have to do with what is invariant from world to world. Now the only things that are invariant from world to world are platonic things – are, in effect, concepts.[60] So if there is a necessary relation between C1 and C2 that must hold in virtue of world-invariant facts, and thus facts about the structures of concepts. And, presumably, it must have to do with the structures of those two concepts specifically. (*x is a triangle* necessitates the truth of *x has more than two sides* because of something about the constitutions of the concepts *triangle* and *two sides.* The same is true *mutatis mutandis* in the case of *x is water* necessitates the truth of *x is* H_2O – even though the relevant facts about the constitution of the concepts *water* and H_2O are not epistemically transparent.)

So given two *object-independent* concepts C1 and C2, no empirical work – no investigation of the spatio-temporal world – is needed to know in what necessary relations they stand with respect to each other. Only purely *conceptual* – purely *a priori* – work is involved. This point, of course, generalizes to any pair of object-independent concepts.

Of course, if one can know a priori in what necessary relations C1 and C2 stand with respect to each other, then one can know what a priori what *possible* relations hold between them.[61]

To sum up, if C1 and C2 are object-independent concepts, then one can know a priori in what necessary relations C1 and C2 stand with respect to each other, and (therefore) in what relations it is possible for them to stand with respect to each other.

But, of course, where object-*de*pendent concepts are concerned, this is not the case. One has to do empirical work to know the constitutions of such concepts; for such concepts have stretches of the empirical world for their

contents, and nothing can be known about the empirical world save through observation (empirical work). So it cannot typically be known a priori in what necessary relations object-dependent concepts stands with respect to each other.

Consider the proposition 'if x is H_2O, then x is water'. This proposition is true in any possible world. Which means that its truth is not contingent on what obtains in this or that world. Its truth is therefore underwritten by something world-invariant; it is, therefore, underwritten by facts about the structure of concepts – specifically, it would seem, the concepts *water* and H_2O. These concepts are object-involving; they have actual spatiotemporal kinds for their contents (or, at least, for part of their contents). Of course, there are many things one cannot know about these kinds without doing empirical work. (One cannot know, without doing empirical work, the facts about those kinds that are relevant to determining whether 'if x is H_2O, then x is water' is true or not.) That is why 'if x is H_2O, then x is water' is not a priori.

In a word, that proposition is a posteriori because the concepts involved – *water* and H_2O – are object-involving; they have for their contents (or, at least, for part of their contents) spatiotemporal kinds.

Let us sum up. Where object-*in*dependent concepts are concerned, logical possibility coincides with actual possibility. Where object-*de*pendent concepts are concerned, this is not the case. I will use this fact as a way of arguing for dualism – as a way of reviving the conceivability argument described a moment ago.

A test of object-independence

It will help if we a have test by which to determine, for any concept C,

whether C is object-dependent or object-independent. It is not hard to produce such a test. In this section I will simply state the test. In the next section I will justify it.

The test is this: If grasping C constitutively[62] involves being casually connected to the external world, then C is object-dependent. If grasping C does *not* constitutively involve being casually connected to the external world, then C is object-*in*dependent.

Let me now justify this. To grasp propositions that are de re about water or Socrates, one has to be causally connected to Socrates or water. So to think *Socrates was wise* or *water quenches thirst*, one has to be causally connected to Socrates or water. Thus, to grasp the concepts *Socrates* or *water* one has to be causally connected to these things. A brain in a vat couldn't grasp these concepts[63], and thus couldn't grasp the propositions *Socrates was wise* or *water quenches thirst.* (In this paper, when I talk about 'brains in vats', I mean brains that are just like the brains of *living*, cognitively normal people, and are subject to the some proximal, but not the same distal, stimulations. So the brains in vats I will be talking about are just like my brain and your brain – except that they are in vats. Of course, in reality, there are brains in vats. But they are dead. They are not what I mean by 'brains in vats.')

In general, to grasp any object-involving concept, one has to be causally connected to something in the non-mental world. A brain in a vat couldn't grasp such concepts.

But a brain in a vat *could* grasp the proposition *2+2=4.* Thus a brain in a vat *could* grasp the concepts *2, =, +* and so on. Surely a brain that was just like yours, but was disembodied, could think *2+2=4.* Would it really be fair to say of a brain just like Gauss's, but disembodied, that it didn't know that 2+2=4?Thus the proposition *2+2=4* is thus object-independent, and so

therefore are the concepts *2*,+,= and so on.

So the way to test whether a concept C is object-dependent or not is this. If grasping C involves having a causal connection to the non-mental world, then C is object-dependent. If grasping C does *not* involve having a causal connection to the non-mental world, then C is object-*in*dependent.

The concept *Socrates* is object-dependent, because to grasp it, one's thoughts have to have a certain causal origin; one has to be connected to the non-mental world in a certain way. (A brain in a vat cannot think about Socrates.) The concept *even number* is object-independent, because to grasp it, it *isn't* necessary that one's thoughts have a certain causal origin or, therefore, that one be connected to the non-mental world in this or that way. (A brain in a vat can do math – can think about the number 2.)

If a brain in a vat can grasp a concept C, then C is object-*in*dependent. If a brain in a vat can*not* grasp C, then C is object-*de*pendent.

The justification for this test

Let me now provide the justification for this test of object-(in)dependence. Let Charlie be some brain in a vat. If Charlie can grasp a proposition P, that means that it is possible to grasp P without having any causal connections to the non-mental world. That, in turn, means that grasping P does not, at least not in principle, have anything to do with what is going on in the world outside one's mind. The state of Charlie's mind *by itself* – without relying on the existence of spatiotemporal things outside that mind -- determines whether P is being grasped or not. Given that fact, it is clear that P's being grasped doesn't depend on there being any spatiotemporal thing *outside* of Charlie's mind. (Of course, if *Charlie's* grasping P doesn't involve the existence of any state of affairs outside Charlie's mind, then *Bob's* grasping it

– and Mary's grasping it and Fred's grasping – doesn't involve the existence of any extra-mental state of affairs.) And given *that* fact, in turn, it is clear that P doesn't have anything spatiotemporal for any of its constituents. So P is object-*in*dependent.

In a word, if grasping P doesn't involve having a causal connection to anything non-mental, then P is object-independent. And so, therefore, are all the concepts which P comprises.

Let us sum up. If whether one grasps C is *not* sensitive to facts about one's causal liaisons to the external world, then C is object-*in*dependent. Otherwise, C is object-*de*pendent.

A proto-argument

Let us say that C1 is some mental concept. (By a 'mental concept' I mean some concept such that, if x falls under it, then x is ipso facto mental: the concept *belief that 2+2=4* is such a concept. By a 'mental concept' I do *not* mean anyone's grasp of anything.) And let's say that C2 is some physical concept. (So C2 is some concept such that, if x falls under it, then x is ipso facto physical. An example of such a concept would be *has a positive electrical charge*.)

If C1 and C2 are object-*in*dependent, and *x falls under C1* is logically compatible with *x does not fall under C2*, then things falling under C1 are *not* identical with things falling under C2. For, to reiterate, where object-*in*dependent concepts are concerned, logical possibility/necessity coincides with *actual* possibility/necessity. So if *x falls under C1* is logically consistent with *x does not fall under C2*, then it *is* possible for things falling under C1 to be distinct from things falling under C2. More formally, for any x, any y, if x

falls under C1 and y falls under C2, it is possible that x is distinct from y. And since identity holds *necessarily*, it follows that x *is* distinct from y.

In what remains of this paper I will show, first, that at least some mental and physical states of affairs can be described entirely in terms of object-*in*dependent concepts; and, second, that the concepts involved are *logically* consistent with some mental entities not being physical entities.

Since, where object-independent concepts are concerned, logical consistency implies actual possibility, it can be inferred that some mental entities really can be distinct from physical entities. And since x *can* be distinct from y only if x *is* distinct from y, it follows that some mental entities are not physical.

The argument proper

Now at least we can put forth our argument for dualism. Suppose the following. Joe and twin-Joe are exactly alike as far as their mental contents *narrowly individuated* are concerned. But Joe and twin-Joe are in utterly different physical environments. Now Joe, like any cognitively normal human being, grasps the concept *belief that 2+2=4.* (In other words, he knows what it is for somebody to believe that 2+2=4.) Given the facts, as we've just described them, does it make any sense to suppose that twin-Joe does *not* have the concept *belief that 2+2=4*?

Surely not. To illustrate this, let us consider the most extreme realization of the facts as we've just described them. Suppose the following. Joe is an ordinary human being on Earth. Twin-Joe is a brain in a vat. But twin-Joe's mental life, narrowly individuated, is just like Joe's. (In other words, twin-Joe's mental life -- considered apart from his being a brain in a vat, and apart

from the all the causal facts associated therewith – is just like Joe's.) Surely twin-Joe, despite his unfortunate predicament, knows just as well as Joe what it is to have the belief that 2+2=4. Surely twin-Joe can manipulate this concept (i.e. the concept *believes that 2+2=4*), and knows its application-conditions, as well as Joe. So we must conclude that the concept *belief that 2+2=4* is object-*independent.*

An exactly analogous argument can be given to show that various other mental concepts – e.g. *desire for a meaningful life, ticklish sensation, love of poetry* – are object-independent. (Some may have misgivings about applying what we said about *belief that 2+2=4* to *is a pain* or *is a ticklish sensation.* These misgivings are unwarranted, and I deal with them below.)

Not *all* mental concepts are object-independent; for some mental concepts implicate object-dependent concepts. For example, the concept *belief that Socrates was smarter than Plato* implicates the object-dependent concepts *Plato* and *Socrates.* So even if two people are exactly alike as far as their mental contents, narrowly individuated, are concerned, it might be the case that one of them has the concept *belief that Socrates was smarter than Plato* while the other does not; for possession of this concept involves, not merely having certain contents, but also having certain causal liaisons to one's environment. Twin—Joe, being a brain in a vat, will not have these causal liasons, and will not grasp this concept, even though Joe does.

But plainly *some* mental concepts are object-independent. The concept *belief that 3 is greater than 2* is object-independent: a brain in a vat could, in principle, grasp this concept as well as anyone. Suppose that twin-Gauss is mentally just like Gauss, except that twin-Gauss is a brain in a vat. Surely twin-Gauss has the same mathematical acumen as Gauss – has the same intelligence about number as Gauss. And surely twin-Gauss is just as able to apply the associated mental concepts – e.g. *believes that there are infinitely*

many primes – as Gauss.

Let us now turn to physical concepts. Obviously some physical concepts are object-dependent, e.g. *identical with water*, *identical with heat.* But it seems to me that the most *fundamental* physical concepts cannot be object-dependent. The purpose of any science is to provide as complete a description of the objects falling in its purview as possible. The purpose of e.g. theoretical physics is to provide as accurate, as fine-grained, and as complete a description of the states of affairs falling in its scope. Now to the extent that a proposition is *object*-involving – i.e. to the extent that it involves object-involving concepts – it has (by definition) actual *objects*, rather than *descriptions* of those objects, for its content. The proposition *Socrates was bald* is object-involving: Socrates himself is a constituent. Inevitably, a proposition that has Socrates himself as a constituent is (ceteris paribus) less information rich than one that contains a *description* of Socrates. If you replace *Socrates*, in the just mentioned proposition, with some description of, say, the mental and physical events associated with Socrates, the resulting proposition will be incomparably more fine-grained, more information-heavy than the former. (I am not saying – what Kripke proved false – that *Socrates was bald* is synonymous with some statement of the form *the unique such and such was bald.* All I am maintaining is that, when objects occurring in propositions are replaced with descriptions, then – holding everything else constant – the resulting proposition, though perhaps not synonymous with the original, is quite obviously much richer in information than the former.) In general, there can be no doubt that, in so far as *objects* are constituents of a proposition, rather than *descriptions* of objects, that proposition is of a lower degree of complexity, and of information-richness, than it would otherwise be.

When objects-occur in propositions, they occur as simples – even though

objects *per se* are not simple. The proposition *Socrates was bald* has a very simple structure -- one of the form *a has phi* – even though Socrates himself was very complex. For whatever reason, Socrates' complexity is not implicated in the proposition *Socrates was bald.* Socrates, when transported into that proposition, somehow becomes a simple. (Why this is so is not clear, but *that* it is so *is* clear.) Somehow, when Socrates becomes a constituent of a proposition, his *internal structure* is not implicated in that proposition. And this point applies to *any* object that becomes a constituent of a proposition. Water is complex; it has a molecular and atomic structure. But this complexity is not implicated in the proposition *water freezes at 32°*; that proposition has a maximally simple form – it has the form *a has phi.* To sum up, whenever a spatio-temporal entity figures as a constituent of a proposition, its *complexity* – its internal structure – is not implicated in that proposition; spatio-temporal objects, in propositional contexts, are utterly *simple.*

But when a *description* occurs in a proposition, all of its complexity *is* implicated in that proposition. So *the greatest bald philosopher of all time was Greek* is more complicated a proposition – and therefore, if true, more information-rich – than *Socrates was bald*; the same being true for any other proposition that results from *Socrates was bald* by replacing *Socrates* with a description. (Once again, I do not deny the following: given only that a proposition P results from *Socrates was bald* by replacing *Socrates* with a description, it doesn't follow that P is synonymous with *Socrates is bald.* In fact, if anything, it seems to follow that P is *not* synonymous with Socrates is bald.)

In general, in so far as a proposition involves object-dependent concepts, it is not as information-rich as it could be.

Now theoretical physics is concerned with generating maximally precise –

maximally information-rich – propositions about sub-atomic phenomena. An idealized theoretical physics really just *is* a set of maximally accurate such propositions. As we've just seen, to the extent that a proposition employs object-dependent concepts, that proposition is not as fine-grained as it would otherwise be. So the foundational concepts of theoretical physics – the concepts in terms of which the most precise and exhaustive description of sub-atomic reality are to be couched – must be object-*independent.* These must not comprise actual chunks of the spatio-temporal world; for in so far as they do, they do not do justice to the internal structure of those chunks.

So the propositions of a physics which captures the fine-grain of the sub-atomic world must be couched in object-independent concepts. The concepts of such a physics must be *object-independent.*

Now let us move on to the next of the argument. As we noted earlier, for a state of affairs to occur in space-time region R is just for some concept to be instantiated in that region. For there to be a particle of such and such charge and mass, moving at such and such velocity in region R, just *is* for the concept *particle with such and such charge and mass moving with such and such velocity* to be instantiated in R – the same being true *mutatis mutandis* for any other state of affairs that might occur in R (or any other region). So for some sub-atomic state of affairs to obtain in R is just for some micro-physical concept to be instantiated in R.

Everyone agrees that whatever physical states of affairs hold in a region R supervenes on what microphysical states of affairs hold in R. The biological, the chemical, the geological, and so on, supervene on the microphysical. This is equivalent to saying: what physical concepts are instantiated in R – whatever biological or chemical or geological…concepts are instantiated in R – supervenes on what microphysical concepts are instantiated in R.

The materialist holds that whatever mental states of affairs are instantiated

in R supervenes on (is strictly determined) by what physical concepts are instantiated in R. This is equivalent to saying: for the materialist, whatever mental concepts are instantiated in R supervenes on what microphysical (atomic and sub-atomic) concepts are instantiated in R. So the materialist holds that some *necessary* relation holds between certain microphysical and mental concepts: some relation of the form *when C1 is instantiated in R, C2 is also instantiated in R*, where C1 is a microphysical concept (e.g. *object with such and such charge.…*) and C2 is a mental concept (e.g. *belief that 2+2=4*).

We've observed that, where object-*in*dependent concepts are concerned, necessary relations among those concepts can be excogitated a priori; since (by the definition of object-independent), such concepts do not have any component of the spatio-temporal world for any of their content, and are therefore to be known through non-empirical – purely conceptual or a priori – labour.

We've also noted that many mental concepts – e.g. *belief that 2+2=4* -- are object-independent, *and* that the foundational concepts of theoretical physics are, or ought to be (ultimately), object-independent.

So if materialism is right – if any belief that 2+2=4 *is* identical with some physical state of affairs – then the following must be true: if the concept *belief that 2+2=4* is instantiated in R, that is in virtue of the fact that some object-independent microphysical concept C (or, more likely, set of such concepts) is instantiated in R. In other words, the instantiating of C in R *necessitates* the instantiating of *belief that 2+2=4* in R. (More plainly, if C is instantiated in R, that necessitates that there be a belief that 2+2=4 in R.)

Further, because C and *belief that 2+2=4* are object-independent, the following holds: the fact that C is instantiated in R would *entail* – would *logically*, not (just) metaphysically, necessitate -- that *belief that 2+2=4* was

also instantiated in R. For recall that, if materialism is right, the truth of *C is instantiated in R* necessitates the truth of *the concept* belief that 2+2=4 *is instantiated in R.* And because the concepts C and *belief that 2+2=4* are object-independent, this necessity is an *entailment*; it is a *logical* necessity – one that can be excogitated a priori.

But I find it very hard to believe that there is any purely logical entailment from *C is instantiated in R*, where C is an object-independent microphysical concept, to *there is a belief that 2+2=4 in R.* It follows that a belief that 2+2=4. It is very hard to believe that the instantiating of some microphysical concept *logically necessitates* the instantiating of the concept *belief that 2+2=4.*

Let us now put all the pieces together and close the argument. Recall that where object-*in*dependent concepts are concerned, necessary relations are the same thing as entailment relations. There is, presumably, no entailment from *C is instantiated in R* to *the concept* belief that 2+2=4 being instantiated *in R.* (In other words, there is no entailment from *C is instantiated in R* to *there is a belief that 2+2=4 in R.*) Both C and *belief that 2+2=4* are object-independent concepts. So the just mentioned lack of entailment coincides with the absence of a necessary connection. So there is no necessary relationship between C's being instantiated in R, on the one hand, and there being a belief that 2+2=4 in R – where, once again, C is any microphysical concept.

Let us sum up what we've said so far. For a state of affairs to obtain in a region R just *is* for a concept to be instantiated in R. So for a microphysical state of affairs to obtain in R just *is* for some microphysical concept to obtain in R. We've established the following: there is no object-independent microphysical concept C such that C's being instantiated in R necessitates there being a belief that 2+2=4 in R. So the truth of any proposition of the

form *C is instantiated in R* -- where C is any object-independent microphysical concept -- is compatible with the falsity of *there is a belief that 2+2=4 in R.*. Since both C and *belief that 2+2=4* are object-independent, this lack of an entailment relation coincides with a lack of *any* necessary relation. It follows that given any microphysical state of affairs S, if S can be described by an object-independent proposition (a proposition that uses only object-independent concepts), then S's obtaining in R does *not* necessitate there being a belief that 2+2=4 in R.

This *by itself* does not prove that a belief that 2+2=4 is not physical. But it *does* prove that such a belief that 2+2=4 is not identical with, or supervenient on, any microphysical state that can be described by object-*in*dependent concepts. From this, however, it is a short step to showing that such a belief is not identical with, or supervenient upon, *any* micro-physical state, or therefore any physical state at all.

Remember what we said earlier. To the extent that a proposition employs an object-*de*pendent concept, it doesn't do justice to the complexity, the structure, of the thing which it concerns. The internal structure of a thing is obliterated when that thing becomes a veritable constituent of a proposition. So any proposition that gives that gives the fine-grain of the microphysical world will use only object-*in*dependent concepts. So *any* microphysical state of affairs, when it is described in its fine-grain, is given by object-*in*dependent propositions, i.e. propositions which use only object-*in*dependent concepts.

A while ago we established the following. For *any* microphysical state of affairs S, if S can be described by an object-independent proposition, then S's obtaining in R does not necessitate that there is a belief that 2+2=4 in R. Since *any* microphysical state of affairs is, ultimately, given by an object-independent proposition, this means that *no* microphysical state of affairs in

a region R necessitates that there be a belief that 2+2=4 in R. This, in turn, means that a belief that 2+2=4 is not *identical with* or *supervenient upon* the occurrence of any microphysical state of affairs in R. So the belief that 2+2=4 is not physical.

Now some analytic functionalists will register the following objection to the argument we just gave:

You say – and your argument essentially presupposes – that there is no entailment from *C is instantiated in R*, where is an object-independent microphysical concept, to *there is a belief that 2+2=4 in R.* But *there is* such an entailment. Given a knowledge of the kinematic and dynamic interrelation of the physical objects in R, it *could* be deduced, quite literally, whether there was a belief that 2+2=4 in R. Consider: let R be the region occupied by some computer. If you knew all the microphysical facts in R – i.e. if you knew just what microphysical concepts were instantiated in R – then you would have as fine-grained a knowledge as possible of the character and organization of the states of affairs obtaining in R. You would, in effect, know everything there was to know about the distribution of mass-energy in R – about the course and intensity of electric currents, the mechanical interactions, and so forth. But it seems to me that, on the basis of this knowledge, you *could* deduce that something in R believed that 2+2=4. You would know that the computer generated such and such and output in response to thus and such input, and you would know the intervening electrical and mechanical facts. Now if we are functionalists about the belief that 2+2=4 – that is, if we say that x qualifies as such a belief *wholly* in virtue of its causal liaisons – then, on the basis of the aforementioned physical facts, one *could* deduce that the computer believed that 2+2=4. So *there would be* an entailment from *C1, C2…Cn is instantiated in R* – where C1, C2…Cn are microphysical concepts

– to *there is a belief that 2+2=4 in R.*[64]

This seems to me to involve a very wrong-headed and overly reductive conception of belief. But I cannot pursue that here. Nonetheless, what the analytic functionalist says about the belief that 2+2=4 has virtually no bearing on examples involving concepts of phenomenally pregnant states – for such states are not plausibly regarded functionally. (It is exceedingly implausible to say that the essence of being a pain or a ticklish sensation is having certain causes or effects.[65]) And we can use this fact to circumvent the objector's point.

Consider the concept *is a pain.* For reasons we've seen, if materialism is right, then the concept *is a pain* is instantiated in R wholly in virtue of the fact that some object-independent microphysical concept C is instantiated in R. Now *is a pain* is object-independent. Suppose that Bob and twin-Bob are exactly alike in respect of their mental states, narrowly individuated (i.e. considered apart from any environmental facts). And suppose that Bob has the concept *is a pain* (i.e. he knows what it is to attribute pain to someone or something). Under those circumstances, could it possibly be *denied* that twin-Bob had the concept *is a pain*? Surely not. Surely if Bob knows what a pain is, then so does twin-Bob: facts about the environmental causes of Bob's and twin-Bob's mental contents are totally irrelevant. If Bob has the concept *is a pain*, and twin-Bob is his exact duplicate – in all respects *modulo* those having to do with the environmental causes of his mental contents – then if Bob grasps the concept *is a pain*, so does twin-Bob.

So *is a pain* is an object-independent concept. And, so by assumption, is C. But surely there is no *entailment* from *C is instantiated in R* to *the concept* is

a pain *is instantiated in R.* For any object-independent microphysical concepts C1, C2…Cn, from the fact that C1, C2…Cn are instantiated in R, it surely does not logically follow that *is a pain* is instantiated in R; there is no *entailment* from *C1, C2…Cn are instantiated in R* to *the concept* is a pain *is instantiated in R.* In other words, there is no *entailment* from *C1, C2…Cn are instantiated in R* to *there is a pain in R.* Now since C1, C2…Cn and *is a pain* are object-independent, necessary relations that hold between the former and the latter are *entailment* (basically, logical) relations. So if, from the fact that C1, C2…Cn were instantiated, it were really *necessary* that *is a pain* be instantiated, there would be an *entailment* from *C1, C2…Cn are instantiated in R* to *the concept* is a pain *is instantiated in R* (i.e. *there is a pain in R*). But there is no entailment; so there is no necessary connection between the concepts *C1, C2…Cn* being instantiated in R, on the one hand, and *is a pain* being in instantiated in R, on the other. (That is, there is no necessary connection between *C1, C2…Cn* being instantiated in R, on the one hand, and there being a pain in R, on the other.) Now C1, C2…Cn stand for *any* object-independent microphysical concepts one might choose. So *for any* object-independent microphysical concepts one might choose, there is no necessary relation between those concepts being instantiated in R, on the one hand, and there being in a pain in R, on the other. For a microphysical state of affairs to obtain in R just *is* for some microphysical concept to be instantiated in R.

So, it follows that *for any* microphysical state of affairs S that can be described in object-independent concepts, S's obtaining in R does *not* necessitate there being a pain in R. From this, of course, it follows that *no* such microphysical state of affairs (in R, or any other region) is identical with the occurrence of some pain in that region.

So if the occurrence of a pain is to be identical with some microphysical state of affairs (and therefore with some physical states of affairs), that

microphysical state of affairs must be identical with one that cannot be described in object-independent terms. But remember what we said earlier. The microstructure of the world must, ultimately, be embodied in object-independent propositions – i.e. in propositions which use only object-independent concepts. For in so far as a proposition uses object-*de*pendent concepts, it does not do justice to the complexity, the structure, of the thing which it concerns. So all maximally precise statements about the microphysical are object-independent. As we've seen, the occurrence of a pain is not necessitated by the occurrence of any microphysical state of affairs that is amenable to object-*in*dependent description. And this means, as we've just seen, that the occurrence of a pain is not necessitated by the occurrence of *any* microphysical state of affairs. And this, in turn, means that pain is not *identical with* or supervenient upon, any microphysical state of affairs; and therefore isn't identical with or supervenient *anything* physical (for everything physical is identical with or supervenient upon the microphysical). So not *everything* mental is physical; materialism is false.

Fodor, Jerry. 1983. *Representations.* Cambridge: M.I.T. Press.

Kripke, Saul. 1980. *Naming and Necessity.* Cambridge: Harvard University Press.

Putnam, Hilary. 1975. "The Meaning of 'Meaning'". In *Collected Philosophical Papers vol.2.* Cambridge: Harvard University Press.

Program-causes and the Inertness of the Mental

I. There are some compelling *prima facie* reasons to think that functionalism and certain forms of externalism strip mental content of any causal powers. If that is the case, then we have a good reason to reject those doctrines. Frank Jackson and Phillip Pettit have produced a brilliant argument purporting to show that, in fact, functionalism and the relevant forms of externalism *are* compatible with the presumption that mental content has causal powers. Their argument involves an original and (in my view) correct point about causation. But I believe that their *application* of this correct point to this particular set of problems is flawed. They have not, I will argue, given us any reason to think functionalism and the relevant forms of externalism *are* in fact compatible with the presumption that mental content has causal powers.

II. It seems reasonable to suppose that, in some cases, brain-states realize mental contents. (I wish the word "state" to be taken as broadly as possible – to apply to both enduring brain-structures and occurrent phenomena.) My belief that 1+1=2 is realized by certain brain-states. My belief that water is wet is realized by a different set of brainstates.

Let x be an arbitrary brain-state of mine. According to functionalism, for x to realize mental content C is for x to have a certain causal role. For x to have a certain causal role is, at least in part, for x to have such and such effects. So x realizes a belief that bread nourishes at least partly in virtue of the fact that x has certain effects – for example, it leads to my choosing to eat bread (in certain contexts).

There is an obvious *prima facie* problem with functionalism. Mental

content is causally efficacious. In any case, this is what we are inclined to think. My believing that bread nourishes is causally effective: it leads me to drive to the store, buy bread, eat said bread, and so on. Given this, suppose that brain-state x has effects E1…En. x's having C presumably has a *causal* hand in bringing about at least some of E1…En. As we just saw, supposing that x is a brain-state realizing my belief that bread nourishes, x's being such a belief is what *causes* me to buy bread, eat it, and so forth. But if x's having those effects is what x's realizing that content *consists in*, then x's realizing that content cannot possibly be what *causes* E1…En to come about. If x's realizing a belief that bread nourishes *consists* in x's causing me to drive to the store and buy bread, eat said bread, and so on, then x's realizing that belief cannot possibly be what *causes* me to drive to the story, to eat the just purchased bread, and so on.

So if functionalism is right, my belief that bread nourishes ends up *not* having a causal hand in my going the store to buy bread; and, by an exactly similar argument, it ends up *not* having a causal hand in my wanting to eat bread, bake bread, and so on. It ends up not having a causal hand in any of the contexts where, quite obviously, it *does* have a causal hand.

The nature of this problem for functionalism becomes clear if we are careful to make a certain distinction. Strictly speaking, it is not *things* that have causal powers. It is not *the rock* which breaks the window, but the rock's moving with a certain velocity, and having a certain mass, and so on. *The rock* doesn't do anything. What does the work is really the rock's having certain properties – a certain mass, solidity, velocity, and so on – along with various other things' having various properties (e.g. the window's having a certain micro-structure).

Not every one of the rock's properties is causally effective in *every* causal transaction involving the rock. Suppose the rock has a certain dimple on one

of its sides. In some contexts, the rock's having that property will be causally effective. (For example, if you pour water down that side of the rock, the presence of the dimple will affect the course of the water.) But in other contexts, the presence of that dimple will *not* be causally effective. (For example, if you drop the rock in a vacuum, the presence of the dimple won't have any effect on the trajectory of the rock.)

Suppose, once again, that x is a brain-state realizing my belief that bread nourishes. Strictly speaking, x doesn't do anything: what does the work is x's having certain properties – e.g. x's having certain electrical, chemical, and cellular properties. Not every causal transaction involving x is one where x's *electrical* properties are causally effective. Perhaps in certain contexts it is x's mass, not its electrical properties, that are causally effective. Similarly, not every causal transaction involving x is one where x's *mental* properties are causally effective. In some contexts, x's mental properties might be causally irrelevant: what is doing the work in those contexts might be x's having a certain electrical charge or its having a certain specific shape. So there is a difference between *x's* being involved in some causal transaction, on the one hand, and *x's realizing a belief about bread* being causally effective in that context. The latter might or might not be causally effective on that occasion.

But, of course, we want to believe that, in at least some cases, my believing that bread nourishes is causally effective. So we want to believe that, in certain contexts, x's *having a certain mental content* is causally effective. What leads me to eat the bread in front of me is my believing that bread nourishes. Here what is effective is not just x, but is x's having the right mental properties.

Here is where the problems for functionalism begin. The functionalist obviously allows that *x* has causal powers (it isn't in the same category as a number or a disembodied spirit). If x's being a belief that bread nourishes is

consists in its having such and such effects – e.g. its leading me to choose to eat bread, to decide to teach my children how to bake bread, and so on – then x's having that content cannot be *causative* of my choosing to eat bread, bake it, teach my children how to bake it, and so on.

Surely x's being a belief that bread nourishes is what *makes* it have a certain causal role: it has a *regulatory* effect on that causal role . x's being such a belief is what *causes* me to eat bread, to buy bread, to teach those I like how to bake bread. But if x's being such a belief is *identical* with its having those effects – with its bringing about those states of affairs – then x's being such a belief cannot be a *cause* of those states of affairs. After all, nothing is a cause of itself. Functionalism doesn't allow mental content to *regulate*, or otherwise effect, x's causal role, since functionalism *identifies* x's having such and such with x's causal role.[66]

Let us sum up. If content is *identified* with causal role, then content cannot be what governs causal role. But in that case, content doesn't *do* anything; it becomes inert, a free-wheel or epiphenomenon. But plainly content is *not* inert; plainly content does a lot. So functionalism must be wrong.

III. Another doctrine – content externalism – also seems to make content inert. Smith and Twin-Smith are atom for atom duplicates. But Smith is thinking about Mary, while twin-Smith is thinking about twin-Mary. The reason is that Smith's brain-state has one distal cause (Mary), while twin-Smith's has another (twin-Mary). The difference in distal cause makes for a difference in content.[67] But this difference in content seems to be inert. What does all the work is *local*: the spatially and temporally local property of

Smith's brain are what cause him to pick up the phone and dial those numbers; the same is true (mutatis mutandis) of Twin-Smith. The differences in distal causes are causally irrelevant: everything that does causal work is in the here and now – is lodged inside Smith's and twin-Smith's respective crania. It is irrelevant that Smith's state was distally caused by Mary, as opposed twin-Mary or hologram-Mary or Robo-Mary. Smith's state might as well have been caused by any one of them, so far as the causal properties of that state are concerned.

But if the externalist is right, then Smith and twin-Smith have qualitatively different mental *contents*, corresponding to the differences in distal causes. Smith is having a thought about Mary, while twin-Smith is having a thought about twin-Mary. So, if externalism is right, that difference in content is inert. [68] What is *not* inert – what does all the causal work – is precisely what Smith and twin-Smith have in common; and what they have in common is neutral between Mary and twin-Mary. The fact that Smith is thinking about Mary – as opposed to twin-Mary or Robo-Mary – is without causal consequences. So far as what he is doing right now, his current condition could have been caused by any one of the three. So to the extent that mental content is to be understood in externalist terms – to the extent that Smith and twin-Smith have qualitatively different mental contents in virtue of their differing causal liasons -- mental content doesn't do anything. Externalism thus appears to strip mental content of causal power.[69]

IV. Jackson and Pettit deal with these problems in a brilliant way. We must distinguish causal *efficacy* from they refer to as causal *relevance.*

In everyday discourse, we say things like the following.

“The liquid in the glass froze (thus expanding); and that is what is caused the glass to shatter.”

The truth is: what caused the glass to shatter is something much more specific; it is the fact that, in certain very specific places, certain very specific pressures were put on the walls of the container.

Consider the statement:

“The presence of a flame under the container caused the liquid to boil.”

Well, what really caused the container to heat up is much more specific: certain specific displacements of energy led to some other specific displacements of energy; and these displacements happened in a very specific region – a region coincident with, or vanishingly close to, the region occupied by the under-surface of the pan. Not every displacement of energy constitutive of the flame caused the increase in temperature. It wasn’t really *the flame* which did the work: it was certain very specific micro-events constitutive of the flame.

One last example:

“The pianist hit a wrong note because *somebody* coughed too loud.”

Well, what *really* caused the pianist to hit a wrong note is much more specific. *Smith* coughed too loud – it wasn’t just *somebody*

Causes are much more specific than our statements typically indicate; and the things we describe as “causes” – *somebody’s* coughing too loud, a *flame’s* being underneath the container, *the temperature’s* being too high – are not really what is doing the causal work; there is always something much more

specific going on; some specific molecule (or atom or quark…) bouncing off of some other molecule (or atom…). In fact, in the end, the only things that are doing causal work are specific, minute displacements of mass-energy. *The flame* is much too expansive and gross an entity to be responsible for the fact that this, that, or any other specific part of the pan heats up; what heats up place x on the pan is always some specific mass-energy displacement.

But when we say things like "*someone's* coughing caused the pianist to hit a wrong note", "the presence of the flame caused the pan to heat up", and "the freezing of the liquid caused the vessel to shatter", what we are saying is surely not altogether wrong. There is plainly a sense in which it is true to say that *somebody's* coughing is what caused the pianist to error, a sense in which it is true to say that the presence of the flame caused the pan to heat up, and so on.

Here is how Jackson and Pettit deal with it. The presence of the flame *programs for* the heating-up of the pan. Put another way, it is *causally relevant* without being *causally efficacious* in the heating of the pan. What does this mean? Given that there is a flame present, it necessarily follows that something which *is* causally efficacious in the heating up of the pan will be present. The flame is constituted by the specific displacements which did the causal work. And the presence of a flame in that place guarantees that *some* such displacements will be present which will do the causal work. So even though the flame *per se* is not what does the work, the presence of the flame is causally *relevant* to the heating up of the pan: the presence of the flame guarantees or "programs for" the presence of something which would do said work.

The same principle (mutatis mutandis) applies in the other cases. What causes the glass to shatter is not the fact that the liquid inside has a certain temperature or volume. It is something more specific: the impact of certain

molecules on certain other molecules. (Actually it is even more specific than that.) But the temperature of the glass is causally *relevant.* Given that the glass has such and such temperature, it is guaranteed, or "programmed for", that there will be the kind of thing which *is* causally efficacious in the shattering of the glass.

What caused Horowitz to hit the wrong note was *Smith's* coughing or *Brown's* coughing: some *specific* person had to cough. The general fact that *somebody or other* in the vicinity coughed isn't what did the trick. But that general fact does guarantee or program for the existence of the sort of singular fact which *would* do the trick. Given that *somebody* or other coughed, it follows that, for some specific x, x coughed: and the fact that x coughed *is* causally efficacious as regards Horowitz's erring. So *somebody or other's coughing* is causally relevant, though not causally efficacious, to Horowitz's erring. It programs for the existence of something which is causally efficacious, even though it does not itself have that property.

Sometimes Jackson and Pettit describe the thing which causally *relevant*, but not causally efficacious, as the "program cause". So the fact that *somebody* coughed is the program-cause of Horowitz's erring; the fact that the flame was underneath the pan was the program cause of the pan's heating up to such and such temperature.[70]

V. Jackson and Pettit say that, if we take into account the distinction between causal efficacy and causal relevance, we can be functionalists or externalists *and* hold onto the idea that content has an important causal role. For the reason earlier given, what is causally *efficacious* in Smith's reaching for the phone and dialing those numbers is not the fact that Smith is thinking about Mary. But Smith's thinking about Mary is *causally relevant.* His doing so "programs for" the kind of thing that will do the trick – that *will* be

causally efficacious in his hitting those seven numbers. So content has causal relevance (though not causal efficacy). So as far what is causally *efficacious* is concerned, it is irrelevant whether the distal cause of his current condition was Mary, twin-Mary, robo-Mary, or a hallucination. Thus Smith's thinking about Mary – as opposed to twin-Mary or robo-Mary – is not causally efficacious. But Smith's thinking about Mary is still *causally* relevant: his thinking about Mary *programs for* his dialing those seven numbers.

The flame *programs for* the specific micro-events E1…En which are causally efficacious in the heating of the pan, even though the flame is not *itself* thus efficacious. By distinguishing causal relevance (the property of programming for something causally efficacious) from causally efficacy, we recover our intuition that the flame has an important *causal* role in the heating of the pan. Similarly, Smith's thinking about Mary *programs for* the occurrences of events E*1…E*n, which are causally efficacious as regards his dialing those seven numbers, even though his thinking about Mary – as opposed to twin-Mary or robo-Mary – is not thus efficacious. Once again, By distinguishing causal relevance from causally efficacy, we recover our intuition that Smith's thinking about Mary has an important causal role in Smith's thinking and acting. In general, by making that distinction, we reconcile our intuitions about the causal importance of representational content in mental and behavioral life with the evident truth of some kind of content externalism.

§ The same principle applies in the case of functionalism. (In this context, let us forget about the problems relating to differences between thinking about water/twin-water, and the like.) For the reasons we described earlier, Smith's wanting water cannot be *causally efficacious* in Smith's reaching for the glass: it cannot, strictly speaking, what causes Smith to do that. But it can

be causally *relevant.* Smith's being thirsty "programs for" his doing that sort of thing; the presence of thirst guarantees the presence of the kinds of things which would to the causal work in Smith's reaching for the glass.

VI. I believe that Jackson and Pettit have had a profound insight into causality and causal explanation. But I don't think that insight helps out functionalism or externalism in the way that they think. First we will deal with the functionalist case.

Let us start with some platitudes. Nothing can be "causally relevant" to itself. A state of affairs cannot be "causally relevant" to its own existence. Obviously things are relevant to themselves. But the relevance isn't *causal.*

Of course, the term "causal relevance" (as Jackson and Pettit are using it) is an invented one; so we can give it any meaning we wish. But surely it would be deeply misleading to say describe the kind of relevance that event E has to itself as *causal.* The relation is better described as logical, or perhaps metaphysical. E is not *causally* related to itself; and it doesn't matter whether we are taking about causal efficacy or causal relevance. Obviously the kind of relation Jackson and Pettit have in mind, when they talk about "causal relevance" is *some* kind of causal relation; it is not a relation which something can have to itself; the word "causal" is there for a reason. So nothing can be "causally relevant" to itself.

The functionalist says that x's causing E1…En is what x's having C *consists in.* There is an identity here, not a causal relation. At the same time, it is clear, from a common-sense point of view, that x's having content C – e.g. x's being a desire for water – has some kind of causal hand in bringing about E1…En (i.e. that it has a causal hand in bringing it about that Smith reaches for the glass). If x's having C is *identical with* x's bringing about E1…En, x's having C cannot be *causally effective* in bringing about E1…En:

after all, x's bringing about E1…En cannot cause x's bringing about E1…En. By the same token, x's having C cannot be *causally relevant* to the bringing about of E1…En. Nothing can be *causally relevant* to itself. If the functionalist is right, then x's bringing about E1…En is *identical* with x's having C. So if the functionalist is right, then x's having C – i.e. its bringing about E1…En – cannot be causally *relevant* to its bringing about E1…En. It cannot be causally relevant, any more than it be causally effective. So the Jackson and Pettit maneuver fails.

VII. To deal with the externalist-case, a background point is in order. Sometimes Jackson and Pettit analyze the notion of causal relevance in counter-factual terms: x is causally relevant to the occurrence of y if, in any counterfactual circumstance where x exists, there also exists something causally effective in bringing about y. The problem is that this is too generous. My striking the glass with the hammer is sufficient, in the counterfactual sense just defined, for the breaking of the glass. The same is true my of striking the glass *while* wearing a yellow hat and having green teeth and wearing a spacesuit…But surely my wearing a yellow hat/space-suit/having green teeth…is not in any sense "causally relevant" to the breaking of the glass. The counterfactual analysis is too broad, too expansive.

The counter-factual analysis of causal relevance needs to be tightened. But we don't want to tighten it too much. For example, we don't want to say that the program-cause is the *smallest* thing x such that x counterfactually guarantees the presence of something causally efficacious. After all, the smallest thing that gives that guarantee is the mass-energy displacement itself. But we don't want to say that *only* that mass-energy displacement is causally relevant: surely we want to allow the flame of which it is a part is causally relevant; we don't want to shrink causal relevance down to causal

efficacy.

The notion of causal relevance is, I believe, to be understood in terms of the notion of *redundancy* – in terms of the notion of a *back-up mechanism.* Consider the case where, as we loosely say, the presence of the *flame* caused the pan to heat up. Let x be the *specific* mass-energy displacements constitutive of the fire which, in the actual case, did the trick. The flame is causally relevant because in any case where you have the flame, but where x (for some reason) didn't do the trick, the flame guarantees that some other mass-energy displacements y *will* do the trick. So, it seems, causal relevance can be understood in terms of "back-up" mechanisms. In any situation where *someone or other* in the vicinity coughed loud enough, but where Smith (the actual cougher) did not cough, that is enough to make Horowitz hit the wrong key.

Now Smith's wearing a yellow hat is *not* causally relevant to Horowitz's erring, because it doesn't constitute the right kind of back-up mechanism; it doesn't amplify the range of cases where something is causally efficacious in getting Horowitz to err. By contrast, the *whole* flame is causally relevant to the heating because, if the flame were smaller, that would eliminate some back-up mechanisms; it would shrink the range of mass-energy displacements that *would* step into the breach in situations where x (the totality of mass-energy displacements that *in fact* did the trick) isn't being causally efficacious. But Smith's wearing a yellow hat, the flame's coming from a green as opposed to a blue lighter – or its coming from a lighter as opposed to a newspaper – doesn't provide this sort of counterfactual reinforcement. It is inert in this respect.

The same is true of the fact that Smith's brain-state has its origin in Mary, as opposed to twin-Mary, or Mary-7, or Robo-Mary, or hologram-Mary or a Cartesian demon or bent light-rays. What is causally *efficacious* is some

congeries X of mass-energy displacements at various contact-points in Smith's nervous system. What is causally *relevant* are the structures that guarantee, in counter-factual situations where X didn't do the trick, some congeries Y would. What is *not* causally relevant is anything that doesn't provide this sort of reinforcement; where that ends, so does causal relevance. Now consider the things going on outside Smith's cranium. Are any of *these* things going to step into the breach in counter-factual situations where X isn't doing the trick? Consider a counter-factual situation where X isn't doing the trick. Is the fact that Smith's brain-state originates with Mary (as opposed to twin-Mary or Mary3) going to help out? Is it going provide needed back-up? No. It is inert. It would provide back-up *only* to the extent that it led to the right kind of *brain-structure* or, in any case, the right kind of *local* phenomenon that would provide back-up.

VIII. In this context, we have to make sure that our terminology isn't doing the thinking for us. Jackson and Pettit are using the term "causally relevant" in a technical sense. But, of course, that expression has a non-technical sense; one could use it in ordinary speech, and be understood. In the non-technical sense, the Big Bang is obviously "causally relevant" to everything that is now the case, including the fact that I'm typing on a computer. In fact, give or take a few niceties relating to quantum-indeterminacy, the Big Bang was causally sufficient for that, and every other, contemporary fact. But the Big Bang is not "causally relevant" in Jackson and Pettit's technical sense.

The concept that Jackson and Pettit are labeling with the expression "causal relevance" is an important, coherent, and explanatorily fecund one[71], and that concept deserves *some* kind of label. But it is easy to conflate the technical meaning of "causally relevant" with its original, non-technical meaning. And such a conflation could easily create the illusion that

the concept of causal relevance has application where it really does not. In fact, I feel that, in their application of their insights into causation to the issues under discussion, Jackson and Pettit themselves may be guilty of this sort of conflation.

Let us use the expression "causally relevant $_{JP}$" to refer to causal relevance in the technical Jackson-Pettit sense. And let us use the expression "causally relevant" to refer to causal relevance in the non-technical, pre-Jackson-Pettit sense. The Big Bang is obviously causally relevant to the fact that I am not typing– after all, it is the ultimate *cause*: and how could it be more causally relevant than that? But the Big Bang is not causally relevant $_{JP}$ to my typing. (Having made this point, let us now go back to using the term "causal relevance" – no subscript – in the technical sense in which Jackson and Pettit use it.)

Now let us consider the paradigms through which Jackson and Pettit give meaning to their technical use of the term "causal relevance". Horowitz hits the wrong note. What causes that to happen is *Smith's* coughing or *Brown's* coughing: it isn't just *someone's* coughing. (Causal efficacy resides in singular facts.) But the general fact that *someone* or other coughed is causally relevant to Horowitz's hitting the wrong note . Why? Because that general fact guarantees that there will be a singular fact of the right kind; it guarantees that, for some individual x, there will be a singular fact of the form *x coughed.* Now what is the nature of this guarantee? We are not talking about a *causal* guarantee. It is not as though *someone's* coughing at time t *causes* Smith's coughing at time t+1, which in turn causes Horowitz to hit the wrong note. No – the guarantee here is much tighter than that. It is in the nature of a logical or metaphysical guarantee. Given that *someone* coughed (in the right vicinity), it follows logically – not causally – that, for some specific x, *x coughed.* So the fact that *someone* coughed logically

guarantees that the right kind of singular fact will obtain.

What causes the pan to heat up is the fact that there are certain *specific* events in an area coincident with, or vanishingly close to, the bottom of the pan. The much more encompassing fact that there is a *flame* underneath the pan isn't what does the trick. What does the trick is much more specific. But the just mentioned more encompassing fact is causally relevant to the heating of the pan. Why? Because it *guarantees* that the right kind of specific events will occur. What is the nature of that guarantee? It isn't a *causal* guarantee. The specific mass-energy displacements which do the trick are *constitutive* of the flame. The flame comprises innumerable mass-energy displacements; and the ones which are causally efficacious in the heating of the pan are among them. The aforementioned guarantee is not *causal.* Here is the wrong story. The flame exists at t, and that causes there to be certain other events E1…En at time t+1; and, in their turn, E1…En cause the pan to heat. That isn't how it works. Here is the right story. The flame is simultaneous with E1…En. This is to be expected, since E1…En are constituents of the flame. The flame guarantees the existence of the right kind of specific pan-heating events in the sense that the flame is *constituted* by all manner events which will do the trick. If the flame were present but, for some reason, E1…En didn't occur, other events E*1…E*n *would* occur, where E*1…E*n would do the trick. It isn't that the flame would *cause* E*1…E*n to happen; it is that the flame would *comprise* those events. The flame is *constituted* by the kinds of things which will be causally efficacious in the heating of the pan. And that is the sense in which the presence of the flame underneath the pan "programs for" the heating of the pan; that is the nature of the guarantee previously mentioned. The presence of the right kind of events is guaranteed not in a causal, but a metaphysical (for lack of a better word) sense: the flame doesn't *cause* there to be the right kind of events; the flame is *constituted by* the right

kind of events.

Exactly similar remarks apply to the case of the liquids shattering the walls of the vessel. Here is wrong picture. The liquid expands at t. That expansion *causes* certain other events E1…En. E1…En, in their turn, cause the walls to crack. Here is the right picture. The liquid expands at t. That expansion *involves* the occurrence of certain events E1…En. And E1…En (the specific mass-energy displacements at, or vanishingly close to, the interior surface of the vessel) cause the vessel to crack. If the liquid-expands, among the events *involved* in that expansion – among the events constitutive of it – are such as will be causally efficacious in the cracking of the glass. So it is not in a *causal* sense that the expanding of the liquid guarantees the existence of causally efficacious events. The guarantee is metaphysical (for lack of a better word): the expansion is *constituted* by the right kind of events.

Let us bring these points to bear on the Mary/twin-Mary case. Smith decides to hit seven numbers on his phone. Mary -- or, better, some state of affairs involving Mary -- is a (contributory) distal cause of this. Let S1 be that state of affairs, that distal (contributory) cause. And let S2 be the event of Smith's hitting those seven numbers. It may be that the occurrence of S1 in some sense "guarantees" the occurrences of S2. But the guarantee here is causal. S1 is buried in the remote recesses of space-time. (To fix our ideas, suppose that Mary and Smith haven't seen each other in sixty years.) There is no sense in which S2 is constitutive of S1 (or vice versa). And, of course, there is no purely logical between them. So the sense in which S1 guarantees S2 isn't remotely comparable to the sense in which *someone's* coughing guarantees that, for some specific x, *x coughs*; and it isn't remotely comparable to the sense in which the presence of the flame guarantees the existence of the right kind of specific mass-energy displacements at, or

vanishingly near, the surface of the pan.

Now *there are* events which, while not themselves being causally efficacious in making Smith dial those seven numbers, do yet *guarantee* that he will do so, where "guarantee" is being used to denote the right logical or metaphysical relationship – the kind that holds between someone's coughing and Smith's coughing, between the flame's being underneath the pan and the pan's being heated. What is it that is causally efficacious in making Smith dial (or at least choose to dial) those numbers? Certain specific mass-energy displacements at, or vanishingly near, certain neural contact points. So what is causally efficacious isn't exactly some chemical (or electrical) storm in Smith's brain – some flurry of z-fibers or whatnot. The *storm* is much too encompassing an entity to be causally efficacious (just as the *flame* is too encompassing an entity to be causally efficacious in heating the pan): causal efficacy resides at (or vanishingly near) contact points; the storm *includes* things that happens at those contact points, but it also includes things that are (relatively speaking) far away from them. Nonetheless, that chemical storm – the thing that neurologists speak of as "the cause" of Smith's number-pushing muscle-spasms – in some sense *guarantees* that the right kind of events will happen at those contact points. What is the nature of that guarantee? The chemical storm *consists* of those events (and others). The occurrence of the storm guarantees the existence of the right events at the contact points because the storm *consists* of events that will do the trick. In the actual world, the storm comprises E1…En – these being the relevant events at the contact-points. But if the storm hadn't comprised specifically E1…En, it *would* have comprised E*1…E*n, where E*1…E*n would have done the trick. It *would* have done this because the chemical storm is composed of the right kind of events – not because it would have *caused* the right kind of events. So what is *causally relevant* to, but not causally efficacious in, Smith's dialing those

numbers is *not* some long gone state of affairs involving Mary: it is some contemporaneous state of affairs which *comprises* the causally efficacious events. Events buried in the past – whether they involve Mary, twin-Mary, or Robo-Mary – are not, in Jackson and Pettit's technical sense of the phrase, "causally relevant" at all. They are *causes*; but they are not causally efficacious – for causal efficacy resides in what is doing the work right now – and they are not "causally relevant" (in the technical sense) because what is causally relevant *comprises* what is doing the work right now.

Notice that, in the paradigms through which Jackson and Pettit define their technical term "causal relevance", the state of affairs that is causally relevant is always contemporaneous with the state of affairs that is causally efficacious. *Someone's* coughing must happen at the same time as Smith's coughing. The flame's being underneath the pan happens at the same time as E1…En occurring at, or vanishingly near, the under-surface of the pan. The liquid's expanding happens at the same time as E*1…E*n happening at, or vanishingly near, the interior surface of the beaker. But with regard to the state of affairs involving Mary – or, better, involving long dead star Beta – in virtue of which I am now thinking about Mary as opposed to twin-Mary, about Beta as opposed to twin-Beta – that state of affairs is quite definitely *not* contemporaneous with the events which are causally efficacious in my dialing those numbers, not even after we take into account the adjustments to our concept of simultaneity demanded by relativity-theory and the like. And this should tell us something as to whether the Mary/twin-Mary case is on all fours with the Smith-someone/flame-E1…En case.

Jackson, Frank; Pettit Philip; Smith, Michael. 2004a. *Mind, Morality, and Explanation.* Oxford: Oxford University Press.

-- 2004b. "In Defense of Folk Psychology." In Jackson, Frank; Pettit Philip; Smith, Michael. *Mind, Morality, and Explanation.* Oxford: Oxford University Press.

-- 2004c. "Causation in the Philosophy of Mind". In Jackson, Frank; Pettit Philip; Smith, Michael. *Mind, Morality, and Explanation.* Oxford: Oxford University Press.

-- "Functionalism and Broad Content". In Jackson, Frank; Pettit Philip; Smith, Michael. *Mind, Morality, and Explanation.* Oxford: Oxford University Press.

-- 2004e. "Program Explanation: A General Perspective". In Jackson, Frank; Pettit Philip; Smith, Michael. *Mind, Morality, and Explanation.* Oxford: Oxford University Press.

-- 2004f. "Structural Explanation in Social Theory". In Jackson, Frank; Pettit Philip; Smith, Michael. *Mind, Morality, and Explanation.* Oxford: Oxford University Press.

-- 2004g. "In Defense of Explanatory Ecumenism". In Jackson, Frank; Pettit

Philip; Smith, Michael. 2004e. *Mind, Morality, and Explanation.* Oxford: Oxford University Press.

Kaplan, David. 1968. “Quantifying In”. Reprinted *The Philosophy of Language.* A.P. Martinich (ed.). Oxford: Oxford University Press.
--1989. ‘Demonstratives’. In *Themes from Kaplan.* Edited by Joseph Almog, John Perry, and Howard Wettstein. Oxford: Oxford University Press.
-- 1989b. “Afterthoughts”. In *Themes from Kaplan.* Edited by Joseph Almog, John Perry, and Howard Wettstein. Oxford: Oxford University Press.

Putnam, Hilary. 1975. “The Meaning of ‘Meaning’”. In *Collected Papers: Volume II.* Cambridge: Harvard University Press.

THE USELESSNESS OF THE RAMSEY-LEWIS SENTENCE

Functionalism may be defined as follows. For something x to qualify as an instance of mental type T just *is* for x to have a certain 'causal role', where having a certain 'causal role' means having certain causes and certain effects. For any mental type T, and any object or state of affairs x, x qualifies as an instance of *wholly* in virtue of having certain causes and certain effects.

So for functionalism, there is *nothing* to being a certain kind of pain, or a certain kind of belief, or a certain kind of intention, besides having certain causes and certain effects. In general, for functionalism, there is nothing to being an instance of a mental type other than having certain causes and effects.

Functionalism is almost always used to abet some form of materialism. A little background will help make it clear why this is so.

For functionalism, what makes x qualify as (say) a belief that snow is white is *not* that x is made of this or that material, but that x is caused by certain things and causes certain other things. And there seems to be no reason why a brain state shouldn't have just those causes and effects. Consider the causes and effects characteristic of a belief that snow is white. There doesn't seem to be any reason why a brain state shouldn't have those causes and those effects. Anything can cause anything; causal connections hold only contingently. So there is no reason why (say) a certain pattern of neuro-chemical activity shouldn't have *just* those causes and effects characteristic of a belief that snow is white. So if *all* there is to being a belief that snow is white is having those characteristics and effects, then there is no reason why a certain pattern of neuro-chemical activity shouldn't be a belief that snow is white.

And therewith the relationship of mind to body vanishes. Since *anything*

with the right causal connections qualifies as (say) a belief that snow is white, and since anything physical could in principle have those connections, there is no problem assimilating beliefs that snow is white to physical states of affairs. The same applies, of course, to any mental type. Thus functionalism successfully reduces the mental to the physical – or, in any case, shows that such a reduction is possible.

But a certain *prima facie* circularity characterizes this picture. Given practically any mental type T, instances of that type have *other mental entities* among their characteristic causes and effects. Consider the belief that *snow is white*. Instances of this type – that is, particular cases of the belief that snow is white – characteristically have certain mental causes and characteristically have certain mental effects. For example, a belief that snow is white might characteristically be caused by certain sense-perceptions, operating in conjunction with certain back-ground concepts and beliefs. And a belief that snow is white might, in conjunction with other background beliefs and concepts, characteristically result in *other* beliefs (e.g. the belief that snow is not green or the belief that snow absorbs less light than tar).

Given *any* mental type, even those whose instances have the most tenuous connections with other mental entities, we *still* find that those instances have *other mental entities* among their characteristic causes and effects. Arguably, even tickles and itches have, among their characteristic effects, other mental states of affairs.

So the attempt to use functionalism to reduce the mental to the physical seems guilty of a certain circularity. Functionalism says that a belief that snow is white will be identical with, or constituted by, a certain pattern of electrical activity *provided* that that electrical activity has the right causes and effects. So *up to this point* functionalism seems to be doing a good job of reducing the belief the electrical activity. But then it turns out that among the

'right' causes and effects are other mental entities. So functionalism, it turns out, reduces the belief that snow is white to electrical activity *plus various other mental entities.* So functionalism reduces the mental to the physical *plus* the mental, and thus doesn't reduce the mental at all. (Cf. Quine's point that conventionalism reduces logic to convention *plus* logic and thus doesn't reduce logic to convention at all.)

David Lewis[72] came up with an argument that -- so it is widely thought – shows functionalism *not* to be guilty of the circularity just described or, if you prefer, that shows the circularity just described *not* to threaten the possibility of using functionalism to reduce the mental to the physical.

I do not believe that Lewis has really succeeded in ridding functionalism of *any* kind of circularity. That is what I will try to show in this paper. He has, I admit, *outlined* a way to 'decircularize' functionalism. But once we scrutinize this outline, we find that, if it were stated in 'non-outline' form, it would itself turn out to be embody a vicious circularity.

Lewis' attempt to de-circularize functionalism makes essential use of the idea that there is a 'Ramsey sentence' for mental phenomena. (We will spend a great deal of time stating just what this means.) Lewis doesn't actually give us that Ramsey sentence itself. He just gives us a recipe for constructing it. But if Lewis' solution is to work, the Ramsey sentence must exist; it must be *possible*, at least in principle, to construct that sentence. If the Ramsey sentence for the mental cannot be constructed – if, in effect, that sentence doesn't exist -- then Lewis' solution fails. I will show, in effect, that there is no way to construct the Ramsey sentence for the mental. The idea that Lewis' solution works is based on tendency look at that solution only at a very abstract and schematic level – a failure not to consider what is really involved in constructing the Ramsey sentence of which Lewis speaks, but which he never actually produces.

II. Let us spend the next few pages stating Lewis' alleged solution. As we know, for a functionalist the nature of any mental type – e.g. the type *pain* – is given by a set of propositions saying what characteristically causes instances of that type and what instances of that type characteristically result in. Imagine that, for each mental type, you have identified the propositions which give the characteristic causes and effects of instances of that mental type. Let P_1, $P_2 \ldots P_n$ be this set of propositions. (So P_1 says e.g. for some specific kind of ache, what the characteristic causes and effects of aches of that kind are.) These propositions may, of course, be conjoined. Let us refer to the conjunction of these propositions as S_1.

A couple of clarificatory points are in order. First of all, S_1 is not a statement about this or that person's psychology; S_1 says *in general* what are the characteristic causes and effects of *types* of mental entities. So the 'mental constants' in S_1 are not going to be anything like 'Buddy's fear of heights' or 'Tom's love of tobacco'; the mental constants will be of the form 'the *type* fear of heights.' So S_1 will (very roughly) have the form 'the characteristic causes and effects of *instances* of the type "fear of heights' are such and such and thus and such.' To sum up, the mental constants in S_1 will refer, not to specific mental entities (not e.g. to Harry's anger at having been given a C on a paper) but to mental types (to the *type* 'anger at having been given a C on a paper').

Closely connected with this last point is another. S_1 does not state that such and such *actual* mental states of affairs obtained. (For – to echo what we said a moment ago – S_1 is not a psycho-biography of Tom or Dick or Harry – of any actual individual.) It describes *possible* sequences of mental states of affairs. More specifically, it says, for each *type* of mental entity, what are the characteristic causes and effects of instances of that type.

It must be pointed out that, for at least some mental types, among the characteristic causes and/or effects of instances of those types are *merely* physical events (or states). (By a *merely* physical entity I mean something that is physical and is not also mental. So e.g. an explosion is *merely* physical because it doesn't mediate mental activity. By contrast, c-fibre stimulation – if it really is identical with pain – is not *merely* physical, although it obviously is physical.) For example the characteristic causes of pain are surely certain merely physical states, e.g. certain kinds of tissue damage, and among the characteristic effects of pain are other merely physical states, e.g. heightened blood pressure, dilation of the pupils. Consequently S_1 is going to contain terms referring to types of merely physical events, as well as terms referring to types of mental events.

It will be useful to stipulate some terminology. Consider some existence claim, say 'for some x, x is green.' Let us suppose that α is a term that denotes something green (e.g. α might be the term 'grass' or 'Kermit the frog'). In that case, let us say that α – or, more strictly, the *sequence* <α> -- 'satisfies' the existence claim 'for some x, x is green.' In general, given any existence claim 'for some x_1, $x_2 \ldots x_n$, $<x_1, x_2 \ldots x_n>$ has property P' (where P is an n-place predicate, i.e. a relation which takes n arguments), let us say of any sequence $<a_1, a_2 \ldots a_n>$ which has property P that it *satisfies* that existence claim. So for example, consider the existence claim 'for some x_1, for some x, x loves y' (equivalently: 'for some x_1, for some x_2 $<x_1, x_2>$ has the two place property *loves*'). The pair <Romeo, Juliet> *satisfies* that existence claim, while the pair <Saddam, Bush> does not.

Another related bit of terminology will be useful. Consider an arbitrary existence claim 'for some x_1, for some $x_2 \ldots$for some xn $<x_1, x_2 \ldots . x_n>$ has property P.' (Again, P must be an n-place property, i.e. a relation which takes n arguments.) And suppose that α is a constant such that, if α replaces x_i, and

the corresponding existential quantifier is dropped (i.e. the 'for some x_i' is dropped), the result is a true claim. In that case, let us say, α *satisfies* the existence claim in question *with respect to the ith variable.* So consider the existence claim 'for some x_1, for some x_2 <x_1, x_2> stands in the relation *loves*' (equivalently: 'for some x_1, for some x_2, x_1 loves x_2'). Now if we substitute the constant *Romeo* for the variable x_1, and drop the corresponding quantifier, the result is 'for some x_2, Romeo loves x_2'; which is obviously true (for there is somebody whom Romeo loves – namely, Juliet). So, as we will say, the constant *Romeo* satisfies this existence claim with respect to the first variable. (Of course, *Romeo* also satisfies that existence claim with respect to the second variable; for Romeo is loved by somebody – namely, Juliet.)

With these bits of terminology in place, let us proceed. S_1 is presumably going to contain a number of occurrences of the term 'pain' (or some synonym: we are assuming for the sake of convenience that S_1 is in English). Suppose that you replace each occurrence of this term with a variable, say x_1. The result will, of course, be an open sentence. Now suppose that you prefix this open sentence with the quantifier 'for some x_1'. The result will, of course, be an existence claim; let S_2 be this existence claim. Now anything which satisfies that existence must have exactly the same causal relations as pain. (This clearly follows from the way in which S_2 is constructed out of S_1, and from the way in which S_1 is constructed out of P_1, P_2…P_n.) So, if functionalism is correct – that is, if something qualifies as pain (or as a belief that snow is white or as a regret at not having majored in pharmacology…) *wholly* in virtue of its causal relations – then *anything* which satisfies S_2 is pain.

Now suppose we do the following with S_2. We take every occurrence of

the term ‘anger’ (or its synonyms) and replace it with a variable x_2. Again, the result is an open sentence. Now prefix this open sentence with the quantifier ‘for some x_2’; and let this new existence claim be S_3. Given how S_3 was formed, it is clear that any sequence which *satisfies* it will consist of a sequence containing a term denoting something which has the same causal relations as a anger and a term denoting something that has the same causal relations as pain, in that order. So if functionalism is right – if something *is* anger or pain wholly by virtue of its causal relations – then any sequence which satisfies S_3 must consist of a term denoting anger and one denting pain, in that order. And – what follows more or less trivially from what we’ve said – anything which satisfies S_3 with respect to the second variable has the same causal relations as anger, and – if functionalism is correct – therefore *is* anger.

Obviously the just described process can be continued until *every* mental term in S_1 has been replaced by a variable and the resulting open sentence pre-fixed by the appropriate existential quantifier. Let S_n be the resulting giant existence claim. Given how S_n was formed, it is clear that *any* sequence $<a_1, a_2 \ldots a_n>$ which satisfies S_n has the following property: for each i, a_i will have the exact same causal relations as the term which x_i replaced. So let us say that S_i was formed by replacing each occurrence of the term ‘belief that snow is white’ (or its synonyms) with x_i, and then pre-fixing the resulting open sentence with the quantifier ‘for some x_i’. The term a_i in the sequence $<a_1, a_2 \ldots a_n>$ will have the same causal relations as the term ‘belief that snow is white’. So if functionalism is correct – if mental things are wholly individuated by their causes and effects – then given *any* sequence $<a_1, a_2 \ldots a_n>$ such that, for each i, a_i has the same causal relations as the term in S_1 which x_i replaced, that sequence will contain each of the mental terms

occurring in S_1. More to the point: S_n will *completely describe* the causal relations of all of the mental terms occurring in S_n; this is because, to reiterate, for each i, a_i has the exact same causal relations as the term in S_1 which x_i replaced.

S_n will, of course, contain no terms referring to physical objects – each of these having been 'Ramsified out' (i.e. having been removed and replaced by a variable, with the resulting open sentence being prefixed by an existential quantifier). So S_n (in effect) describes the causal relations characteristic of mental types (at any rate, those mentioned in S_1) *entirely in terms of physical inputs and outputs.* At the same time, S_n does justice to the fact that mental entities are individuated (at least in part) by their relations to *other* mental entities.

So, to sum up: (i) S_n gives a complete statement of the causal relations characteristic of all the mental entities referred to in S_1; (ii) it describes these relations entirely in terms of physical inputs and outputs (i.e. without using any terms denoting mental entities); and (iii) it does justice to the fact mental entities are individuated (at least in part) by their relations to *other mental entities.* Condition (iii) is important because a functionalism that does not recognize that mental entities have constitutive ties to *other* mental entities collapses into behaviorism or, in any case, some kind of eliminativism about the mental.

This sums up Lewis' way of making functionalism be non-circular – of reducing (in a functionalist manner) the mental to the physical, as opposed to reducing the mental to the physical *plus* the mental. The essence of Lewis' solution, it should be apparent, is to describe the causal relations characteristic of the class of mental entities *as a whole* in terms of inputs and outputs: *this* can be done wholly in terms of *physical* inputs and outputs. By contrast, any attempt to describe the causal relations of a *particular* kind of

mental state of affairs in terms of physical inputs and outputs – e.g. the type pain – will necessarily be a description of that state of affairs in terms of physical inputs and outputs *and* mental inputs and outputs.

III. Recall how the Lewis sentence is formed. For each mental type M_i, we consider the sentence (or proposition) that says what the characteristic causes and effects are of instances of M_i; and we let P_i be that sentence. We then *conjoin* P_1, $P_2 \ldots P_n$; and we let S_1 be this sentence. Finally, we 'Ramsify out' the mental terms in S_1.

There is a major problem with this analysis. Any given mental type can be caused by infinitely many different types of states of affairs and can result in infinitely many different types of states of affairs. (I will defend this point in a moment.) What I will now try to show is that, because this is the case, the sentence that Lewis describes either (depending how you look at it) doesn't exist or must contain unreduced mental terms.

Consider the belief that snow is white. There are infinitely many ways a person could arrive at this belief. A person could arrive at it by seeing white snow. Or a person could arrive at it by taking somebody else's word for it. Or, most importantly, a person could arrive at it through some kind of scientific inference. Imagine a race of people so small that they couldn't actually *see* that snow is white: they could nonetheless figure out that snow is white by inferring what the reflective properties of snow are through some complicated theoretical apparatus. Plainly there are, in principle, *infinitely* many belief systems – infinitely many deductive or cognitive machineries -- through which one could figure out, or otherwise ascertain, that snow is white.

And just as plainly, the belief that snow is white could feed into infinitely many different belief systems and thus have infinitely many different effects.

If you are forced to learn that snow is white through some elaborate body of physical theory – because, say, you are too small (or too big) to *see* that snow is white – then obviously the repercussions of your learning that snow is white will be very different from the repercussions of, say, an ordinary sized human's learning that snow is white. To sum up, there is no limit to the variety of different ways one could learn that snow is white, and no limit to the variety of different effects having this belief might have within the economy of one's intellect. In a word, there are *infinitely* many causal roles that this belief could play.

Obviously someone who wishes to state the Ramsey-Lewis sentence cannot individually mention *each* of these causal roles. Remember, the functionalist says that for x to be a belief that snow is white just *is* for x to play a certain causal role. Now if you then ask the functionalist, 'what is the causal role played by the belief that snow is white?', he *cannot* say 'well, it is the causal role played by the belief that snow is white': obviously that would not be a satisfactory answer – at least from *his* point of view. Remember that someone trying to construct the Ramsey-Lewis sentence for the mental is trying, in effect, to *eliminate* mental terms (like 'belief that snow is white') in favor of statement about the causes and effects of instances of the mental type corresponding to that term. So terms like 'belief that snow is white', 'delight in the fact that snow is white', etc. cannot figure in a completed Ramsey-Lewis sentence; they must all be Ramsified out. So someone trying to construct the Ramsey sentence for the mental has to *get rid of* terms like 'belief that snow is white', in favor of mention of the *causal roles* of instances of the mental types associated with those terms. So if you ask someone trying to construct the Ramsey sentence for the mental 'what is the causal role played by the belief that snow is white?', his answer cannot be 'well, it is just that - -the causal role played by the belief that snow is white.'

That cannot be his final answer because, if it were, then his version of the Ramsey sentence would contain an 'unreduced' mental term (namely 'belief that snow is white' or some synonym), and it is the main purpose of the Ramsey-Lewis to get rid of such terms. No – he must specify the causal role of the belief that snow is white *without* using the term like 'belief that snow is white' or any cognate term.

To put this point another way: suppose you say to the functionalist 'look, there are infinitely many causal roles that could be played by the belief that snow is white: what do all these causal roles have in common?' The functionalist *cannot* answer this by saying 'what they all have in common is that they are all occupied by the belief that snow is white.' He must be able to say what that causal role is *without* having to define that causal role in terms of the belief that snow is white. For if he cannot do this, then he will be left with an 'unreduced' mental term, and his Ramsey sentence for the mental will not do what it is supposed to – it will not eliminate mental terms and thus, *a fortiori*, will not eliminate them in favor of non-mental terms.

But mental types are typically such that their instances can be caused in *infinitely many different ways* and are such that their instances can have *infinitely many different effects.* There are *infinitely many different ways* that the belief that snow is white could be acquired, and there are *infinitely many different effects* that such a belief could have. So there are *infinitely many different causal roles that such a belief could play.* A belief that snow is white can be located on infinitely many different causal paths.

Given that there are infinitely many causal roles that a belief that snow is white could have, the person tasked with giving the Lewis sentence now has a problem; for he obviously cannot mention each of these causal roles individually: he cannot just give a *list* of all these causal roles. So, plainly, he must mention something that *all* of these different causal roles have in

common; he must identify some property that *all* (and only) these causal roles have in common. There is, of course, something that all and only these causal roles have in common: they are all causal roles occupied by the belief that *snow is white.* But remember what we said a moment ago: the functionalist must be able to state what these various roles have in common *other* than their all being occupied by the belief that snow is white. (For if he does not do this, then his Ramsey sentence will contain an 'unreduced' mental term, and will thus not do its job of eliminating the use of such terms.) So let us ask: *apart* from all being occupied by the belief that snow is white, what do these infinitely many causal roles have in common? Equivalently, *without using a term like 'belief that snow is white' (or a cognate term*), is there some way to state what all these causal roles have in common.

I very much doubt that that there is any property that all and only these causal roles have in common *apart* from being occupied by the belief that snow is white. Obviously the causal role of such a belief will be determined (at least in part) by what kinds of inferences people make from that belief, and also what kinds of inferences lead them to form that belief. So (assuming that sentient beings have some minimal tendency to be rational), the causal role of the belief that snow is white will 'track' the inferential role of the proposition *snow is white.* Now think about all the inferences that involve this belief – that this belief 'validates'. *Jerry thinks that grass is the color of snow*; *so Jerry infers that grass is white. Zeldar thinks that tennis ball are either the color of snow or are round-squares; so Zeldar thinks that tennis balls are white.* I very much doubt that there is some property that all and only these transitions have *other* than there being validated by the belief that snow is white. (More exactly, I doubt that there is some property that all these transitions have in common *other* than there being validated by the belief that snow is white *and other* than their being validated by beliefs that

contain or involve commitment to this belief – e.g. the belief that *either* snow is white or 2=3 or the belief that if snow is not white, then the Pythagorean theorem doesn't hold for Euclidean triangles. If we say that what the aforementioned transitions have in common is that they are validated by one of these exotic equivalents of the belief that snow is white, then we are still stuck with an unreduced mental term: for we are still identifying the class of all these transitions by saying that what all its members have in common is that they are validated by the *belief* that e.g. either snow is white or 2=3.)

To sum up, any given belief can occupy *infinitely* many causal roles. In fact, as a matter of empirical fact, the same belief (i.e. belief in the same proposition) occupies different causal roles in any two people you might choose. The functionalist has to say what all these causal roles have in common *other* than their being occupied by the belief in question. In any case, the Lewis sentence *eliminates* mention of 'the belief that snow is white' in favor of mention of its causal role. So anyone trying to give the Lewis sentence must mention what all these causal roles have in common other than being occupied by the belief that snow is white. There doesn't seem to be any way to do this. The belief that snow is white can validate infinitely many different transitions. There is no way – at least no way I can see -- to identify this class of transitions except as the class of all and only those inferences validated by the belief that snow is white. This class is, of course, infinitely large; so its members cannot be listed. So this class must be identified by citing some *property* that all its members have in common. There *is* such a property, of course: all its members have the property that they are validated by the belief that snow is white. But someone trying to construct the Ramsey-Lewis sentence for the mental cannot identify this class in that way. For reasons we stated earlier, in order to avoid having an unreduced –

un-'Ramsified out' – mental term in his Ramsey-Lewis sentence for the mental, one must identify some property that all members of this class have in common *other* than that they are validated by the belief that snow is white (or trivial variants of this belief, e.g. grass is either green or not green). And, it is safe to say, there is no such property: the only way to identify to say what objects (events, states of affairs) fall into this class is *in terms of* the belief that snow is white (or in terms of some equivalent belief, e.g. *if snow is not white, then 2 is greater than 4*). In any case, I challenge anyone to produce some other way of singling out the membership of this class. [1]

Stephen Stich makes some points very similar to those made in the last few pages:

> Typically, if a person believes that everyone aboard the morning flight to Bamako was killed in a crash, and if he comes to believe that the prime minister of Mali was on that flight, those two beliefs together will cause him to believe that the prime minister is dead. Typically, if a person wants to go to New Zealand, and if she comes to believe that she must get a visa in order to make the trip, she will acquire a desire to get a visa. Endlessly many further examples might easily be constructed. But therein lies a problem, for there is no hyperbole in the prospect of *endlessly many* [Stich's emphasis] further examples. There are literally infinitely many inferential paths leading both to and from every belief…If a person believes [that Niger is west of Mali] and comes to belief that if Niger is west of Mali then Niger has more than two million inhabitants, he will typically come to believe that Niger has more than two million inhabitants. And so on, ad infinitum. Moreover, obviously, this is just the beginning. (Suppose he believes that if Niger is west of Mali, then Niger grows 5 percent more grain, or that if Niger is west of Mali, then Mali is more politically stable, or…[Stich's ellipsis]) If we are

to construct our implicit functional definition of 'believes that Niger is west of Mali' from the conjunction of these commonsense platitudes about inference, then the definition will have to be *infinitely long!* [Stich's emphasis.] Now some stout-hearted souls are not intimidated by infinitely long conjunctions, provided they can be recursively summarized….But it is far from clear that the class of inferences recognized by common sense *is* [Stich's emphasis] recursively characterizable.[73]

What I have been maintaining is that the list of inferences – the list of inferences leading both to and from (say) a belief that *Niger is west of Mali* is (i) infinite and (ii) *not* capable of any recursive characterization– *except* a characterization that *employs* the term 'belief that Niger is west of Mali' (or some synonym) or some analytically equivalent mental term (perhaps an analytically equivalent term would be. '*pro-attitude* towards the proposition *Niger is west of* Mali'). The point is that, unless we help ourselves to the very mental terms we are trying to define – in this case *belief that Niger is west of Mali* – we cannot recursively characterize the causal role of that belief (and therefore cannot define it at all: for a non-recursive characterization of it would be an infinitely long, and therefore unstateable, conjunction). At the same time, if we *do* help ourselves to that mental term, or some analytic equivalent, then our characterization will obviously not be appropriately *eliminative*; we will be stuck with an unreduced mental term and will have failed to state the causal roles of instances of mental types wholly in terms of physical entities.

I would like to discharge a bold assertion I made a moment ago. Once again consider the type *belief that snow is white*. Let C be the class of characteristic causes and effects of instances of that type. I said:

The only way to say what objects fall into this class is *in terms of* the belief that snow is white (or in terms of some equivalent belief, e.g. *if snow is not white, then 2 is greater than 4*). In any case, I challenge anyone to produce some other way of singling out the membership of this class.

I would like to substantiate this arrogant claim. For I think that, as a matter of logic, it is provable that the challenge in question cannot be met.

As we've noted, the functional role of the belief that snow is white tends to track its inferential role. In any case, any *functionalist* will want to say that this is the case. (Maybe certain kinds of non-functionalists could deny it.) It makes little sense, for a functionalist, to say that the functional role of a belief is totally divorced from its inferential role. After all, a functionalist individuates beliefs by their functional roles; so if a belief doesn't have the functional role that it would be inferentially appropriate for the belief that snow is white to have, then *ipso facto* that is good reason – at least for the functionalist – *not* to see that belief as the belief that snow is white. So, to sum up this leg of the argument, for the functionalist, it is nearly enough analytic that the functional role of a belief should, to some degree of approximation at least, track the inferential role of the proposition it bears. Put another way, for the functionalist, a belief that snow is white should at least *tend* to track the inferential role of the proposition that snow is white.

Now consider the class of all ordered pairs of propositions <P, Q> such that given P and *given also* the proposition that snow is white, Q can be inferred. There are infinitely many such pairs, of course. And as a matter of logic, there is no way to identify this class except *in terms of* either the proposition *snow is white* or some logically equivalent proposition (e.g. *either snow is white or 4 is an odd number*). In other words, there is no way to give the membership of this class except by saying that it consists of all

pairs <P, Q> such that given P and given either the proposition *snow is white* or some logically equivalent proposition, Q follows. For the functionalist, as we've noted, the functional role of the belief that snow is white *tends* to track the inferential role of the proposition *snow is white*. In fact, for the functionalist, it is *definitive* of a belief that snow is white that it *tend* to track the inferential role of *snow is white.* So, for the functionalist, the functional role of the belief that snow is must be understood (at least in part) as a tendency to track this inferential role. And there is no way to define this inferential role *except* as that which is occupied by the proposition *snow is white* (and propositions logically equivalent thereto, e.g. *either snow is white or there are square circles*). So, for the functionalist, the belief that snow is white just *is* (inter alia) something which tracks a certain inferential role; and there is no eliminative definition of this inferential role – that role can be stated only by stating the proposition which occupies that role. So, in the Ramsey-Lewis sentence, there will be *ineliminable* mention of the proposition *snow is white*. And the definition of the *belief* that snow is white which the Ramsey-Lewis sentence provides will include some clause like: *that entity whose causal role tracks the inferential role of the proposition snow is white.* And (at least arguably) there is no way to state that causal role *except* in terms of 'tracking the inferential role proposition that snow is white' (or some cognate expression): there is no way to describe the belief that snow is white except in terms of the property of tracking the inferential role of the proposition that snow is white, and there is no way to describe this inferential role except in terms of the proposition snow is white – there is no 'eliminative' description of that role. So the Ramsey-Lewis sentence is *irreducibly* committed to describing the belief that snow is white as something that (inter alia) 'tracks the inferential role of the proposition that *snow is white*'.

But – to close this argument -- to speak of a spatiotemporal entity (and mental entities are in this category) as *tracking* the *inferential role* of a proposition: to speak of an entity in those terms is, it seems to me, to speak of it in *mentalistic* terms. Why is it mentalistic, exactly? First of all, it is a description that makes use of a *normative* notion – the notion of a *correct* inference. So in describing something as 'tracking a certain inferential role', we are describing it in terms of *correctness* conditions – in terms of its conforming to some logical or normative standard. Now when we describe a spatio-temporal process as being correct or incorrect, right or wrong, valid or invalid – when we describe it normative terms – it seems we are describing it in *mental* terms.[74] Non-mental spatio-temporal things aren't right or wrong, valid or invalid (except in a derivative sense[75]). Mental things are right or wrong, valid or invalid. So if there is no way to give the causal role of the belief that snow is white except in terms of its being something that tracks a certain inferential role – that is, if there is no way to state this causal role except in *normative* terms – then it seems to me that there is no non-mentalistic description of that causal role. Now if we could cash out that causal role in, say, purely neural or bio-chemical terms, then we would have a non-mentalistic description of it. But, as we've seen, we cannot identify that causal role except in terms of its being the role of something that 'tracks' a certain 'inferential role': there is no non-normative way to identify that causal role. Now a 'normative' description of some spatio-temporal process is a mentalistic one. And if that normative description cannot be cashed out in terms of some non-normative one – e.g. in terms of one using only norm-neutral concepts like *neural interaction*, *electric current* etc. – then that description is *irreducibly* mentalistic. Now we've seen that the Ramsey-Lewis sentence will *ineliminably* contain norm-heavy terms – e.g. terms like

‘tracks the inferential role of the proposition *snow is white* – for, as we’ve also seen, there is no way to describe the causal role of the belief that snow is white without using such ‘norm-heavy’ terms. Since irreducibly norm-heavy descriptions of spatio-temporal entities are irreducibly mentalistic descriptions of them, it follows that the Ramsey-Lewis sentence will ineliminably and irreducibly contain mentalistic terms. And this means that it is a failure, since the sole purpose of that sentence is to give an entirely *non*-mentalistic description of the mental – one that uses only the non-normative, purely descriptive terms of *bona fide* physicalistic inquiry: terms like ‘synaptic transmission’, ‘neural pathway’, etc.

To put this more clearly, the purpose the Ramsey-Lewis sentence is to reduce the mental to the physical. The Ramsey-Lewis sentence does this only to the extent that it gives the causal roles of mental entities in purely non-mental terms. But we’ve seen that it gives those causal roles, at least in part, in ineliminably mentalistic terms. So it *doesn’t* succeed in reducing the mental to the physical. This isn’t to say that the mental *doesn’t* in some sense reduce to the physical – I think it does. It is to say that the Ramsey-Lewis sentence doesn’t give this reduction. And since the Ramsey-Lewis sentence is, as far as I can tell, the best way of eliminating *functionalist* forms of materialism from circularity, we have a case here that functionalist forms of materialism *are* in fact guilty of the circularity of which (as we saw earlier) they are *prima facie* guilty. Which would suggest that the reduction of the mental to the physical is to be carried out along non-functionalist lines (supposing, what there is some reason to think is the case, that such a reduction can be done).

I should consider one more objection:

You say that if something spatio-temporal is describable in normative

terms, then it is *ipso facto* mental. But this is wrong. Consider the sentence token ‘Socrates is bald’. To describe that thing as a ‘sentence token’ is to describe it as having truth-conditions, meaning etc. So it is to describe it in normative terms. But that thing is completely non-mental – it is just a pattern of ink on a page.

If the token “Socrates was bald” was *just* a pattern of ink on a page, then it *wouldn’t* have truth-conditions, meaning etc. In a world where there were no sentient beings, that entity *would* in fact be just a pattern of ink on a page – and in such a world it would have no normative dimension. This means that what has a normative dimension – what has semantics etc. – is not the pattern of ink *per se*, but is rather the pattern of ink *as* something that is associated with various intentions, beliefs etc. To put it another way, what is meaningful, and what has a normative dimension, is not, strictly speaking, the pattern of ink by itself, but something like the ordered pair of that pattern of ink and various mental entities which, in some way or other, assign it meaning. But that ordered pair obviously *is* something mental, at least in part. So with regard to the objector’s point that something altogether non-mental can have a normative dimension, that point is simply false.

Just to be clear on this: what has a normative dimension is not the pattern of ink “Socrates is bald” *per se*; for if that thing *per se* had a normative dimension, it would have that dimension in a world devoid of sentient beings. But in such a world it would have no normative dimension; it would just be dead ink. So the existence of sentient beings is a conceptually necessary condition for that thing’s having a normative dimension. And this means, in effect, that what has a normative dimension is not that thing, taken by itself, but some complex – some ordered pair or sequence, perhaps – which includes

some mental things (presumably, beliefs, linguistic intentions, linguistic practices etc). And *that* complex is not altogether non-mental, as it includes mental things (e.g. linguistic intentions, apprehensions of propositions etc). So the objector's point that altogether non-mental things can have a normative dimension is superficial (it is based on the superficial idea that a sentence is *just* a deposit of ink/stream of noise) and false.

I will end by making a point that will forestall a possible confusion concerning the structure of our argument. It is patently obvious that a finite sentence, or finite list of finitely long sentences, can generate an *infinite* list of sentences or, perhaps, can generate an infinitely long sentence. (I can say: consider the sentence formed by adding "and snow is white" infinitely many times to the sentence "grass is green". Here, at least arguably, I have given a finite rule for constructing an infinitely long sentence.) Any axiomatization of a discipline (e.g. Hilbert's axiomatization of geometry) consists of a finite number of sentences/propositions that generate finitely many such propositions.

The problem with the sentence we have been calling the Ramsey-Lewis sentence is *not* that it *generates* an infinitely long sentence or that it *generates* infinitely many sentences. (Remember, the Ramsey-Lewis sentence is the sentence which Lewis describes whose purpose is to de-circularize functionalism.) That wouldn't be a problem at all; it would be a virtue. The problem with the Ramsey-Lewis sentence lies in two points. First, it is that it *itself* is infinitely long. Second, there is no *rule* by which the Ramsey-Lewis sentence can be constructed *except* for some rule which contains terms which it is the Ramsey-Lewis sentence's whole purpose to *eliminate.* So *either* the Ramsey-Lewis sentence is an infinitely long sentence such that there is no *rule* for constructing it. *Or* the Ramsey-Lewis sentence is a finite sentence which ineliminably contains the terms which it is the Ramsey-Lewis

sentence's purpose to eliminate, and thus irremediably committed to the existence of the things which it is the Ramsey-Lewis sentence's purpose to eliminate.

Of course, some might say:

> I don't see anything wrong with the idea of a sentence such that (i) it is infinitely long and (ii) there is no rule for constructing it. So the Ramsey-Lewis sentence is one of those sentences. I don't see what is so wrong with that. Frank Ramsey said that there were infinitely long series of numbers such that there was no rule for generating those series.[76] So why couldn't there be infinitely long sentences that cannot be generated by finite rules?

First of all, what Ramsey said about numbers is highly controversial. Second, even if it weren't controversial, it isn't clear whether it applies to propositions or sentences. The idea of a *sentence* or *proposition* that is infinitely long but that cannot be generated by a finite rule is hard to make sense of. A proposition is the kind of thing that is grasped; there is, I should think, an internal connection between the concept of being a proposition and the concept of being graspable. But something that is infinitely long and cannot be generated by a finite, graspable rule could not itself be grasped.

But, most importantly, even if we do concede that there are infinitely long propositions that cannot be generated by a finite rule – even if we concede that there are ungraspable propositions[77] – Lewis' solution would still fail. Lewis point is this:

> We can show that functionalism is not circular. Here how you do it. You do

it by means of the following sentence [which we have called the Ramsey-Lewis sentence]. Here's how you construct the Ramsey-Lewis sentence; in other words, the Ramsey-Lewis sentence is the thing you get by following *these* rules.

So Lewis' whole solution goes through only if the Ramsey-Lewis sentence can be produced or, in any case, identified in some way or other. Lewis hasn't *actually* given us the Ramsey-Lewis sentence. So Lewis hasn't given us any proposition at all – not one that is finitely long, not one that is infinitely long. He's only given us a recipe for constructing this proposition. But what we have seen that *either* there is no way to apply this recipe at all *or,* if we *can* apply this recipe, the resulting sentence *contains* the very terms it is supposed to eliminate. For the recipe tells us that the Ramsey-Lewis sentence is formed out infinitely long lists; for the Ramsey-Lewis sentence is formed by taking *all* the causes/effects of instance of the type *belief that snow is white* – all infinitely many of them – and replacing *each* occurrence with a variable…But an infinitely long list things cannot be generated, except by rules. And a sentence cannot contain an infinitely long list; it can contain, at most, a rule for generating that list. So the Ramsey-Lewis sentence contains rules for *generating* lists of all the causes/effects of instance of the type *belief that snow is white.* The Ramsey-Lewis sentence cannot itself contain mention of each individual cause/effect. But we've seen that these rules *themselves* ineliminably contain the terms that the Ramsey-Lewis sentence is supposed to eliminate. So the Ramsey-Lewis sentence is formed out of rules which *ineliminably* contain terms like *belief that snow is white, desire for world peace,* and so on. But the whole *raison d'etre* of the Ramsey-Lewis sentence is to *eliminate* such terms. So if the Ramsey-Lewis sentence *can* be

constructed, it doesn't' do its job (which is to eliminate terms like *belief that snow is white*). But to say that the Ramsey-Lewis sentence cannot be constructed is just another way of saying that the Ramsey-Lewis sentence doesn't exist. So Lewis' solution is a failure.

Intentionalism

Abstract

Recently, many have argued that phenomenal content *supervenes* on representational content; i.e. that the phenomenal character of an experience is wholly determined (metaphysically, not causally) by the representational content of that experience. This paper does two things. First, it identifies many counter-examples to intentionalism. Second, it shows that, if intentionalism were correct, that would require that some extreme form of representational atomism also be correct: a form of atomism that is widely held to have been refuted by Sellars (among others). An argument, not found in the existing literature, is set forth to the effect that a sufficiently extreme form of representationalism atomism is not feasible. Various intentionalist counter-ploys are considered and neutralized.

Introduction 1-3
Part I: Some Counter-examples to intentionalism 3-9
Part II. A theoretical argument against intentionalism 9-32
An outline of our argument 11-12
Why some form of conceptual non-atomism is true 13-17
Why conceptual non-atomism entails that intentionalism is false: a prima facie case 17-19
Why phenomenal content doesn't supervene on propositional content 19-22
Why the phenomenal content of sensations doesn't supervene on non-propositional content (supposing that there is such a thing) 23-26
An argument against the existence of non-propositional content 26-32

Introduction

The sole purpose of the present inquiry is to make a case against a doctrine often referred to as "intentionalism." Some background will help to make it clear what intentionalism is.

Consider an arbitrary perception – say, a perception of a brightly colored object. There seem to be two quite distinct senses in which this perception has 'content'. On the one hand, this perception *represents* the world as being a certain way. So, as we might say, the perception has 'representational content.' On the other hand, there is, to use Nagel's expression, 'something it is like' to have that perception. So, as we might say, the perception has 'phenomenal content'. Alternatively, it has a 'phenomenology'. (In this paper, the expressions 'such and such has phenomenal content' and 'such and such has a phenomenology' will be used interchangeably.)

Not every mental entity that has representational content has phenomenal content. For example, I believe that 2+2=4. But there is not 'something it is like' to have this belief (though, of course, there is something it is like to experience the conscious manifestations of that belief). All 'non-occurrent' or 'dispositional' mental entities lack phenomenal content (so it would seem, at any rate, and so I will assume). At the same time, if cognitive scientists are right, some occurrent mental processes lack phenomenal content. Consider the sub-personal computational processes mediating vision or language comprehension, if such exist: presumably these are not phenomenally pregnant, even though they are occurrent.

What seems clear is that conscious experiences have phenomenal content. Seeing a sunset, tasting burnt toast, imagining a funeral -- there is something

it is like to have such experiences.

The question arises: given a mental entity that has both phenomenal and representational content, what is the relationship between these two dimensions of that mental entity? What is the relationship between the *representational content* of e.g. seeing a sunset, on the one hand, and the *phenomenal content* of that experience, on the other hand? Is this relationship purely contingent? Could an experience with that exact representational content have a utterly different phenomenal content? Or does a difference in the former *necessarily* involve a difference in the latter?

A doctrine sometimes referred to as 'intentionalism' provides a clear answer to these questions: according to intentionalism, the phenomenal content of an experience is *wholly determined* by its representational content. What is meant by this is *not* that the representational content of an experience *causally* determines what phenomenal content it has. What is meant is that it is impossible in the strictest sense of the word – *metaphysically* impossible – that two experiences should have exactly the same representational content but differ phenomenal content: just as it is impossible that two statues should have the exact same physical properties but different aesthetic properties; and just as it is impossible that two creatures should consist of atoms and sub-atomic particles interrelated in the exact same way but differ in their chemical and biological properties.

Consequently, for intentionalism, it is metaphysically impossible that two experiences should have the same representational content, but different phenomenal content -- there cannot be representational identity alongside phenomenal diversity. To embody this in the current jargon: for intentionalism, phenomenal content 'supervenes'[78] on representational content.[79] [80]

The Purpose and Structure of the Present Inquiry

This purpose of this paper is to argue *against* intentionalism. The structure of this paper is basically this. First, I will show that there are a number of apparent counter-examples to intentionalism; there are, by all appearances, many cases where the same bit of representational content is associated with different types of phenomenal content. This happens in Part 1.

Part 2 gives a very different argument against intentionalism. There I will try to show that phenomenal content supervenes on representational content only if an extremely implausible doctrine that I will refer to as 'representational atomism' is true. So part 2 gives a kind of conditional refutation of intentionalism: *if* conceptual atomism is false, then so is intentionalism. And I will give some prima facie reasons to think that, indeed, conceptual atomism is false.

Part 1 Some counter-examples to intentionalism

In this section, I will identify some apparent counter-examples to intentionalism, and I will try to discredit ways of neutralizing these counter-examples.

To begin with, perceptions associated with different modalities can obviously overlap (possibly even coincide) in representational content, while differing dramatically (perhaps completely) in respect of phenomenal content. I can smell the rottenness of the fish and I can also taste it: here we seem to have identity of representational content and diversity phenomenal

content. I can feel the roundness of the ball and I can also see it: identity (or at least overlap) of representational content without a corresponding overlap in phenomenal content.

There is another *prima facie* problem for intentionalism. Let us begin by considering the proposition: "The representational content of an experience *wholly* determines its phenomenal content." Given this, suppose that experience E has representational content R and phenomenal content P. If E's phenomenal content is *wholly* determined by E's representational content, that means, presumably, that there is nothing about E *other than its having R* in virtue of which E has P. So, it would seem to follow, having R is a *sufficient condition* for having P. From this, in turn, it very much seems to follow that given *any* mental entity, if that entity is to have P, it is *enough* that it have R. (If we deny that having R is, in general, enough for having P, then we are denying that the *one thing* about M that enables it to have P is that it has R; and we are therefore denying that M has P *wholly* because it has R. But intentionalism just *is* the position that phenomenal content is *wholly determined* – not *partially* determined – by representational content.) So to say that, where *experiences* are concerned, phenomenal content is wholly determined by representational content – to say this is to say that, if an experience has phenomenal content P and representational content R, that means that having R is *sufficient* for having P; and this, in turn, means that *anything* – whether an experience or not – that has R must also have P. So intentionalism – which I have defined as the view that two *experiences* cannot differ in representational content without differing in phenomenal content – entails the view that *no* two entities can differ in representational content without differing in phenomenal content.

Once this is seen, a rather striking shortcoming on the part of intentionalism becomes evident. Ordinarily it is thought that the

representational contents of perceptions can be *stored* (in their entirety) in non-occurrent memories; so that a non-occurrent memory sometimes (at least in principle) has the same representational content as a perception. But, of course, non-occurrent memories have *no* phenomenal content, while perceptions *do* have phenomenal content. So, vacuously, non-occurrent memories do not have the *same* phenomenal content as perceptions, even though (presumably) they may have the same representational content as perceptions. But – if what we said a paragraph ago is correct – intentionalism is (at least implicitly) committed to the view that no two mental entities can differ in phenomenal content while coinciding in representational content.

These are not, I grant, insuperable problems for the intentionalist. For example, the intentionalist may *deny* that non-occurrent can have *quite* the same representational content as perceptions. He may say that, although a perception can certainly overlap in representational content with a non-occurrent memory, there is never *identity* of representational content; and this margin of representational difference accounts for the total disparity in phenomenal content.

This is not, in my view, a very promising response. Even if it is granted that non-occurrent memories do not have *quite* the same representational content as perceptions, they obviously can overlap in representational content to an extremely high degree; so we'd expect there to be a corresponding phenomenal overlap. Also, the bits of representational content that *can* be stored in non-occurrent memory are oftentimes strongly associated with phenomenal contents when those bits of information are embodied in perceptions. For example, you can have a non-occurrent memory of a colour or a smell or a taste; and obviously these bits of representational content, when they occur in perceptions, subserve phenomenal contents: so why don't they subserve phenomenal content when they are embodied in non-occurrent

memories? In any case, it is strange that non-occurrent memories never have *any* phenomenal overlap with perceptions, given to what a high degree the former may overlap with the latter in terms of representational content.

The intentionalist may try to neutralize our second objection in a similar way:

To be sure, one can feel the roundness of the ball and one can also see it; and yet there is no apparent phenomenal overlap between these perceptions. But any visual perception of the ball will have representational content not had by a tactile representation of it, and vice versa. (When you see the ball, you don't just see its shape; you see its colour, its distance from you etc.; but you can't really feel these things. When you feel the ball, you don't just feel its shape; you feel its firmness, its temperature etc.; but you can't really see these things.)

But – to echo our last counter-objection – granting that visual perceptions don't (fully) coincide in representational content with tactile perceptions, it seems problematic for intentionalism that my visual perception of the roundness of the ball should have *no* phenomenal overlap with my tactile perception of the roundness of the ball. In fact, if you focus on the bit of phenomenal content that, in your visual perception, mediates your perception of the roundness of the ball; and if you focus on the bit of phenomenal content that, in your tactile perception, mediates your perception of the roundness of the ball; you find that even *these* have no phenomenal overlap. This is, presumably, no small problem for intentionalism.

Before moving, I'd like to consider a couple more attempts to neutralize the foregoing:

You said that non-occurrent memories can have the same representational content as perceptions, while having no phenomenal overlap with them – it being the case that non-occurrent memories have *no* phenomenal content, while perceptions obviously do. But maybe non-occurrent memories are really just dispositions to have certain conscious experiences (of course, all conscious experiences have phenomenal content). So, if that is the case, then – at least after a fashion – non-occurrent memories *do* have phenomenal content.[81]

To begin with, there is one technical objection to this point of view. If non-occurrent memories are, as the objector says, mere *dispositions*, then strictly speaking they have no phenomenal content; for a disposition *per se* doesn't have phenomenal content, even if it is a disposition to have certain phenomenal contents. But, quite possibly, this is a mere technicality, which doesn't really speak to the larger point the objector is making. So I do concede that *if* non-occurrent memories are dispositions to have conscious experiences, this may well undermine one of the arguments I gave against intentionalism.

The objector's move is, from a dialectical standpoint, a good enough one; for I cannot really *refute* the conception of memory that he puts forth. What I can say is that this move commits the objector to what is demonstrably some highly dubious psychology, one that *prima facie* does a serious injustice to the psychodynamics of non-conscious mental activity. If psychoanalysis and cognitive science are to be given any credence (some say they aren't), then a great deal of mental activity happens at an unconscious level and that there is more to being an non-conscious mental entity than being a disposition to produce certain conscious mental entities. You go to sleep remembering P

and also believing if P then Q. But, priori to going to sleep, you haven't 'inferentially integrated' these two contents; you haven't inferred Q. But you wake up believing Q. The obvious explanation is that, while you were unconscious, you (or some component of your mind) inferred Q. Obviously the contents of your recollection that P and your belief that P implies Q were implicated in this operation. So the obvious explanation is that a grasp of representational content – e.g. a non-occurrent memory -- can be operative at an unconscious level. Now there is no reason why your recollection that P, and your belief that if P then Q, shouldn't heavily overlap in representational content with an occurrent perception. Your recollection that P might be your recollection that a certain apple looked like *that* (where the *that* corresponds to some visual presentation – say, of an apple that is firm and green and healthy looking). And your belief that P then Q might be your belief that, if an apple looks like *that* on Monday (nice and firm and green), then it looks like *that* (brown and soft and rotten) a month later. Basically, there is good reason to believe that representational contents can be integrated *at the sub-conscious level* – that *qua* item not falling within the purview of consciousness, a mental content can be operative. And there is also every reason to believe that such entities can overlap heavily with phenomenally pregnant conscious perceptions. (Leaving aside the niceties in the foregoing: Obviously we sometimes have mental 'records' of appearances of objects, and such records *per force* overlap in representational content with phenomenally pregnant sense-perceptions. And it is very hard to believe that there is *nothing* to such records besides their being dispositions to have certain conscious experiences; if 80 years of psychology are to be given any credence, they have a little more autonomy that than allows. And, plainly, such records are *of* things – e.g. colors, smells – that are represented in consciousness by phenomenally pregnant conditions.)

Now, as I said earlier, I do concede that the objector has identified an intentionalist counter-attack that would be hard to *refute*. But, it seems to me, that counter-attack is embedded in some controversial, if not downright implausible, psychology.

The next intentionalist response has, I think, even less force against the foregoing anti-intentionalist arguments (though, as we will see later, it has some force against other anti-intentionalist arguments):

Some philosophers of mind distinguish between propositional and non-propositional representational content.[82] [83] [84] The former is the kind of representational content that can be embodied in a proposition; the latter, the kind of representational content that cannot be so embodied. It seems that the only kind of content you've spoken of so far is propositional content. As you point out, a visual perception of a round ball and a tactile perception have no phenomenal overlap but a lot of representational overlap. Now both perceptions share a certain propositional content, viz. *the ball is round*. And, as you've pointed out, the two perceptions do not have any phenomenal content corresponding to that representational content. But maybe what the intentionalist should say is that, while phenomenal content doesn't supervene on *propositional* content, it supervenes on *non*-propositional content.

Of course[85], one thing *we* could say, by way of a counter-response, is that the distinction between propositional and non-propositional content is spurious (this is what MacDowell[86] and Sellars[87] would say). Of course, such a position is much too controversial to take for granted (though I will

argue for it later on[88]). So, if only for the sake of argument, I will take it for granted that there is a valid distinction here, and will evaluate the objector's proposal accordingly.

It is, I think, pretty clear that two perceptions that are associated with *different* sensory modalities, and that don't have *any* phenomenal overlap, can have a significant degree of representational overlap. The example above shows this. Surely I can *feel* the roundness of the ball, and I can also *see* it: and if I were a bat I could, in effect, *hear* the roundness (I could determine through echo-location that the ball was round); and if I were a mole, I could, in effect, *smell* the roundness of the ball. So, in principle, the representational contents of, respectively, an auditory, an olfactory, a visual, and a tactile perception could overlap. But I very much doubt that any two of these perceptions would have any phenomenal overlap. So regardless of whether the content of a perception is 'propositional' or 'non-propositional' – 'conceptual' or 'non-conceptual', 'analogue' or 'digital' – we *do* seem to have a case of representational overlap that is not matched by any corresponding phenomenal overlap. And this fact is surely threatening to intentionalism. So even if the objector is right in distinguishing between propositional and non-propositional content – and I leave it open for now that he may be – his point doesn't have much force against our argument; for his objection leaves intact *our* point that there can be representational overlap without a corresponding phenomenal overlap.

We have identified what seem to be fairly obvious problems for intentionalism. Strangely, in the now extensive pro-intentionalist literature, none of these problems for intentionalism has even been recognized, let alone dealt with. And that is why the objections we considered were made by a hypothetical interlocutor, and not by actual authors.

Consider, for example, Alex Byrne's important article 'Intentionalism

Defended' (Byrne [1]), which is the most recent major defence of intentionalism. Byrne doesn't consider *any* of the problems cited above: he doesn't consider the fact that (e.g.) the rottenness of the fish can be given to me through smell *or* taste, and thus through different phenomenal contents; he doesn't consider the fact that the roundness of the ball can be given to me through touch *or* sight, and thus through different phenomenal contents. In fact, Byrne's argument embodies a complete disregard for all of these problems.

Byrne's argument is essentially this. Focus on some object, e.g. a blue chip. (This is Byrne's own example.) Your cognizance of the blue of the chip involves your having a certain kind of phenomenal content -- a certain kind of *visual* phenomenal content. Now if there is a change in that particular bit of content, then there is a change in how the chip is represented to you – more specifically, in what colour the chip is represented to you as being.[89]

The same thing is true *mutatis mutandis* of any perception. Suppose that you are tasting a drop of honey, i.e. that you are having a 'gustatory perception' of a drop of honey. Now suppose that the phenomenology of your gustatory perception changes. In that case there is *ipso facto* in the taste that the drop of honey is represented to you as having.

In general, given any perception, there can be no change in the phenomenal content of that perception without a concomitant change in its representational content. Consequently, the representational supervenes on the phenomenal, i.e. there cannot be a variation in the phenomenal content of an experience cannot unless there is a variation in the representational content of that experience.[90]

Byrne's argument is a formidable one: but the conclusion is vastly overdrawn. This argument shows at most[***91***] that where *perceptions*

belonging to the same modality are concerned, there cannot be a difference in phenomenal content without a difference in representational content. But there seems to be no straightforward way to adapt this argument to show that e.g. given perceptions belonging to *different* modalities, or given a perception and a non-occurrent memory there cannot be a difference in phenomenal content without a difference in representational content. So as against Byrne's argument one can say this:

There *are* cases where there is difference in phenomenal content without a different representational content I need only consider a visual perception of the roundness of the ball and a tactile perception of the roundness of the ball, and note that the two perceptions have no phenomenal overlap (at least none corresponding to their representational overlap); I need only consider a perception and a non-occurrent memory of that perception, and note that the two have no phenomenal overlap, despite identity, or near identity, of representational content.

Of course, these putative counter-examples to Byrne's argument are not necessarily decisive; but they certainly have *prima facie* force. And Byrne doesn't even address them. In failing to do this, Byrne is by no means alone: *no* intentionalist, to my knowledge, has dealt with these problems, or even acknowledged them.

Part II. A theoretical argument against intentionalism

So far we've attacked intentionalism by identifying some apparent counter-examples to it. I would now like to propose a very different kind of argument against intentionalism: one that is not based on putative counter-examples to

it, but on considerations of a deeper kind. I'd like to set forth what might be described as a *theoretical* reason to hold that phenomenal content is not identical with, and doesn't supervene upon, representational content.

More specifically, I will try to show that, given the truth of a doctrine we will call 'representational non-atomism', it is simply not feasible to hold intentionalism in any form. By 'representational non-atomism' I mean the doctrine that any ability to represent something to oneself (e.g. the ability to represent things to oneself as being a certain color or as having a certain shape or pitch) presupposes the ability to represent other, related things to oneself. In effect, there are no 'foundational' representational capacities.

The structure of my argument must be understood in terms of a distinction that we talked about earlier. We noted earlier that, according to some philosophers, there is such a thing as *non*-propositional content: in other words, not *all* representational content can be embodied in a proposition. One of the main motivations for this view is as follows. A proposition is essentially a 'digital' representation; it is neatly decomposable into minimal parts. The sentence 'Mary kissed Bob' seems to be neatly decomposable into 'Mary' , 'Bob', 'kiss' plus the tense-marker '-ed'. And presumably the proposition expressed by that sentence is capable of some kind of similar decomposition. Now a sense-perception (or a photograph) doesn't seem to be capable of this kind of decomposition. Given this, some people have inferred that the representational content of (e.g.) a visual perception cannot be embodied in its entirety in a proposition (or set of propositions). Consequently, the representational content of (e.g.) a visual perception is, at least in part, 'non-propositional'.

As I said earlier, some philosophers (e.g. MacDowell, Sellars) think that *all* representational content is propositional. But this position is hotly debated; and I will not take its truth for granted here.

Given this, my argument will have the following form. First, I will argue for 'representational non-atomism.' I will also make it more clear just what I mean by that term.

Then, I will argue that, because representational non-atomism is true, phenomenal content cannot supervene on the kind of representational content that is embodied in propositions.

I will then consider the proposal that there might be some other kind of representational content – that there is such a thing as 'non-propositional' content -- and that phenomenal content might supervene on this other kind of content. I will try to show that the truth of representational non-atomism is every bit as incompatible with the idea that representational content supervenes on non-propositional content (supposing, for the sake of argument, that there is such a thing) as it is with the idea that representational content supervenes on propositional content.

Finally, I will outline some reason to hold that, indeed, *all* representational content can be embodied in propositions – that, in effect, there is no such thing as non-propositional content. If this is correct, then obviously the intentionalist cannot say that phenomenal content supervenes on non-propositional content. (So, in effect, we will provide *two* arguments against the idea that phenomenal content might supervene on non-propositional content.)

An outline of our argument

There are different ways in which we can be conscious of things; different ways in which we can have things in mind. I can have a sensory perception of John. I can have a mental image of John. And I can have a belief about John. With regard to any ability to have a certain thing or kind of thing in mind, let

us refer to this ability as a *representational capacity.* So in virtue of being able to have John in mind, I have a certain representational capacity.

And with regard to one's ability to experience a certain kind of phenomenal content – to have an experience that has a certain phenomenology -- let us refer to this as a 'phenomenal capacity'. So consider the experience of tasting of vanilla ice-cream; consider what it is like to have such an experience. In virtue of your having the ability to have an experience with that phenomenology, you have a certain 'phenomenal capacity'.

A terminological point is in order. I will often talk about a creature's ability to 'represent' things as having some property; I will often say things like 'a creature's ability to represent things as triangular presupposes an ability to represent things as being closed.' For a creature to 'represent' x as having property P is, as I am using these terms, merely for it grasp things – either perceptually or cognitively or imagistically – *as* having P. So for a creature to 'represent' x to itself as triangular is for it to see it as triangular, or to think about it as triangular, or to imagine it as triangular, and so on.

When I say that a creature can 'represent' things as being (say) triangular or as being a certain shade of pink, I don't mean necessarily that that creature can grasp the property of triangularity or of that shade of pink *in the abstract* – I don't mean that it necessarily has concepts of triangularity or of that shade of pink *in isolation* of particular objects that are triangular or have that shade of pink. As I am using the term 'represent', if a creature can *see* things as being a certain shade of pink, or as being triangular, then *ipso facto* it can 'represent' the properties of triangularity and of pinkness. (Of course, if a creature can grasp the concepts of triangularity and pinkness in the abstract, then *ipso facto* it can 'represent' the concepts of triangularity and pinkness. But the converse may well not be true.) So, to sum up, when I say that a creature can 'represent' things as having a certain shade or as having a certain

shape, I *don't* mean necessarily that it can grasp the concepts of that shade or that shape in the abstract – I don't mean that it can necessarily cognitively isolate the concepts of that shape or of that shade. I mean only that it either *see* things *as* having that shape or shade, or that it can imagine things as having that shape or shade, that it can think about things as having that shape or shade, and so on.

Given how we defined the term 'representational capacity', for a creature to be able to 'represent' things as having a certain property is the same thing as for that creature to have a certain 'representational capacity.'

Now if intentionalism is correct, then there is obviously a very close connection between our representational and our phenomenal capacities. To be more precise, supposing that intentionalism is correct, then whatever phenomenal capacities we have, we have these *entirely in virtue of having representational capacities;* and it is impossible that a creature should be able to have a phenomenal capacity *without* having some representational capacity. It is clear why representational is committed to this. If intentionalism is correct, then whatever phenomenal content a mental state has, it has it wholly in virtue of having a certain kind of representational content. Consequently, it would not be possible for one to be able have a mental content with a certain phenomenal content unless one were able to have a mental content with a certain representational content.

But reflection exposes what appears to be a massive difference between representational capacities, on the one hand, and phenomenal capacities, on the other: a difference that is incompatible with the claims of intentionalism. Recently, probably beginning with Wilfred Sellars, philosophical and psychological work has stressed the *non-atomistic character* of our representational capacities: basically, any one representational capacity presupposes a whole slew of others; one cannot have a single representational

capacity without being in possession of a vast and articulated network of them.[92] Now it is extremely hard to believe that one's capacity to feel woozy or to have an itch involves such a network: it seems, at least intuitively, that one could have the capacity to feel woozy without being in possession of a vast network of representational capacities. *If* this is correct, then one's phenomenal capacities are *independent*, at least in some measure, of one's representational capacities. Consequently, one could have mental contents with at least certain kinds of phenomenal content that altogether lacked any kind of representational content.

In what remains of this paper, I will defend the controversial assumptions made in the argument just sketched -- in particular, the assumption that representational capacities come in networks, and not in isolation – and I will consider multiple objections that might be made to this argument.

Why some form of conceptual non-atomism is true

Let us begin by defending the conceptual non-atomism just mentioned. To begin with, one cannot *just* see an object: one must see it *as* having this or that property; as being round or square or blue. And one cannot *just* think about, or imagine, an object. One must think about it/see it/imagine it *as* round or square or sweet. In general, one cannot *just* represent an object to oneself; one must represent it as having these or those properties.

Now it seems that if one can represent one property to oneself, then one can ipso facto represent others to oneself: each representational capacity has 'constitutive' connections to other representational capacities. For example, consider the property of triangularity. If one can represent something as triangular – if one can see/think about/ imagine/…something as triangular – then *ipso facto* one can represent things as having other properties: as e.g.

being straight, being planar, being closed. And in order to grasp any of these properties, one must be able to grasp others. For example, if one can think about the property of being closed, one can think about the property of geometrical continuity (for a closed figure is one with a continuous periphery); and if one can think about this property, then presumably one can think about others; for the property of continuity is presumably not a simple one. (Exactly what one must grasp to grasp the property of continuity is a delicate question. But it seems unlikely that one could grasp this property in a vacuum. Presumably the ability to represent things as continuous presuppose the ability to represent things as having spatial location and a spatial arrangement. And, by all appearances, the abilities needed to represent things as having these properties involve possession of other representational capacities: perhaps e.g. these abilities presupposes the ability to represent things as persistent which, in its turn, involves the ability to represent things as having temporal properties, and so on.)

What we just said about *thinking* about triangularity is true of *seeing* (or otherwise sense-perceiving) things as being triangular. Surely if one can see things as being triangular, then one must also be able to see things as being straight: for one couldn't see any objects as being triangular unless one could see some objects as being straight – one must see straight lines to see a triangle. And one must be able to see figures as being closed and planar: for one couldn't see things as being triangular unless one could see things as being flat or closed – one must see flat and closed things to see triangular things. So the ability to perceive things as triangular presupposes the ability to see things as having other properties.

To sum up, it is pretty clear that one could not represent things as triangular unless one had various other representational capacities.

I submit that this point can be generalized to *all* properties: given *any*

property P, if one is to be able to represent things as having property P, then one must be able to represent things to oneself as having properties other than P. How is this to be shown? Well, first of all, all properties are individuated –- are constituted -- at least in part, by their relations to other properties. And it seems impossible that one should be able to represent things as having some property unless it was also able to represent things as having at least some of the properties constitutively related to P.

Permit me to elucidate this. It is an essential fact about the property of weighing 300 lbs that, if something has that weight, it weighs more than 200 lbs and less than a 1000 lbs. It is an essential or individuating fact about the property of being square that is something has that property, it has the property of being rectangular and does not have the property of being round. It is a essential fact about the property of being pink that if something is pink it is closer in colour to something which is red than to something which is green. These points hold *mutatis mutandis* for any property one might think of. Each property exists in a 'space' of other properties; and, I suggest, one can no more represent things as having a given property without being able to locate that property in a space of other properties than one could represent a given spatial co-ordinate without being able to represent a space comprising other co-ordinates.

This is *obviously* the case with degree-properties: properties like that of weighing a thousand pounds or even the property of simply being heavy. For such properties are clearly relational; and the information one acquires in learning of a thing that it has such a property has a *patently* contrastive or relational character. (Surely one couldn't represent things as weighing 1000 lbs unless one could represent things as weighing more than a 1000lbs and as weighing less than a thousand lbs: it is totally inconceivable that I should be able to entertain the idea that someone might weigh 1000 lbs unless I should

also be able to entertain the idea that one might weigh less or more than a 1000 lbs. And – what is connected with this -- the information that one obtains in learning of a thing that it is 1000 lbs is obviously contrastive: in learning a thing a thing weighs 1000 lbs one is learning, inter alia, that it weighs *more* than one pound.)

The question is whether this point (mutatis mutandis) holds of *all* properties. For example, is the property of being a certain shade of pink in any sense a 'degree'-property? Is it, in any relevant way, like the property of weighing 1000 lbs? Or could one represent things as having this property without being able to represent things as having any other properties? I now wish to show that what is obviously true of one's ability to represent things as having degree properties -- and, more generally, as having *obviously* relational properties -- is true of one's ability to represent things as having *any* property, even those that seem quite non-relational.

For the sake of argument, suppose that there were a creature that could represent things as being some specific shade of pink but that could not represent a thing as having any other colour. (Let P be the just mentioned specific of shade of pink.) Obviously that creature could not, even in principle, represent things as having any colour besides P. So, from that creature's viewpoint, *everything* would either have colour P or would not have any colour at all. So that creature could not *distinguish* the colour of a thing having P from a thing having some other colour. That creature couldn't even *imagine* an object with a colour other than P – for the ability to do that obviously presupposes the ability to represent things as having some colour other than P. So, from that creature's viewpoint, *everything* – everything perceptible, imaginable, thinkable – would either have *the same* colour or would have *no* colour. So that creature could not represent to itself what is *distinctive* about having P – could not represent what makes a thing having P

different, in respect of colour, from a thing with some other completely different colour. But obviously such a creature simply *couldn't* represent things as having P: being able to represent things as having P obviously presupposes being to represent how with that colour are *different* (in respect of colour) from things with other colors -- presupposes being able to represent what is *distinctive* about things with that colour. Surely if one cannot represent what is *different* about things with P, then one cannot represent things as having P at all.

To sum up, being able to represent what is *distinctive* about things with a certain colour presupposes being able to represent how things with that colour differ, in respect of colour, from things with other colors. And being able to do the latter obviously presupposes being able to represent things as having those other colors.

The following analogy might help clarify the structure of the argument just given. Suppose someone had a concept of what it was to weigh 323 lbs, but not of what it was to have any other weight. In that case, that creature's concept of having *any weight at all* would coincide with its concept of having a weight of 323 lbs. So that creature could not possibly know how a creature weighing 776 lbs was different, in respect of weight, from one weighing 323 lbs (or 4 lbs). But if one doesn't know how a creature weighing 323 lbs is different from one weighing 321 lbs or 776 lbs, then surely it can't really be said to know what it is to weigh *specifically* 323 lbs

To sum, it makes no sense to suppose that a creature could represent things as having P, but not as having any other colour: being able to represent things as having some *specific* colour involves being able the difference (in respect of colour) between things with that colour and things with other colors: and being able to do the latter obviously involves being able to represent things as having different colors. So being able to represent a

specific colour presupposes the ability to represent things as having other colors.

An exact analogue of this argument can be given for *any* property: the property of being a square, of being an A-flat, or of having the timbre of a French-horn. To take one of these examples: suppose that a creature could represent sounds as having the pitch *A-flat* but couldn't represent sounds as having any other pitch. In that case, it couldn't represent what is *special* or distinctive about sounds having that pitch; it couldn't represent how sounds that have that pitch differ from sounds that are some other pitch – say, a D-sharp. For if it is to be able to do any of these things, a creature must be able to represent different pitches: and it is just this that, by hypothesis, the creature in question cannot do. Now if a creature cannot hear, imagine, conceive of the difference in pitch between *any* two pitches, then surely it cannot represent things as having some *specific* pitch – as being (say) an A-flat. Being able to represent sounds as having the pitch *A-flat* presupposes being able to represent what is *distinctive* about sounds with that pitch – presupposes being able to represent what makes sounds with that pitch *different* from sounds with some other pitch. And it is just this that the creature in question cannot do. So with regard to the idea that a creature could represent things as having the pitch *A-flat*, but not as having any other pitch – that idea is self-defeating. So being able to represent sounds as having some specific pitch presupposes being able to represent sounds as having other pitches.

Basically, to be able to represent things as having a certain property, one must be able to represent what is *distinctive* about things with that property. And the ability to do that latter presupposes being able to represent things with *other* properties.

There is more to say on this matter. For a creature to be able to represent

(e.g.) a specific shade of pink, it is not necessary *merely* that it be able to represent other colours to itself. It would seem that to represent the *specific* character of (e.g.) P, one must be able to represent *how* P differs from other colours and shades. Surely a creature that cannot represent *how* things having P are different from things having other colors *ipso facto* simply cannot represent *the* difference (in respect of colour) between things with P and things of other colors: being able to represent the difference (in respect of colour) between things that have P and things that have some other colour is surely the same thing as being able to represent *how* things with P differ from things with other colors. And the ability to represent *how* things with P differ from things with other colors would seem to require possession not just of a multiplicity of representational capacities, but of an *organized* multiplicity of representational capacities: of a network of such capacities. Knowing *how* red things differ from pink things and from green things seems to involve knowing how the 'color-distance' between red things and pink things compares with that between (say) red things and green things. So knowing *how* red differs from pink seems to involve having some kind of a map of a system of color-differences, and thus seems to presuppose possession of a *network* of representational capacities.

To sum up, being able to represent things as being P presuppose being able to represent *how* things with P differ from things with other colors. And being able to do this seems to involve having an organized network of representational capacities. So being able to represent things as having P involves having an organized network of representational capacities.

Let us sum up. There is some reason to believe that representational capacities have a non-atomistic character; that if one can represent one thing or property to oneself, that is at least partly in virtue of his being able to represent other things or properties to himself: so if one is to have a single

representational capacity, one must have a whole network of them.

Although this view is quite popular, I have tried to give a defence of it. My defence may or may not have been credible. But the position defended is, I think, credible independently of my defence and is also very widely held. [93]

Why conceptual non-atomism entails that intentionalism is false: a prima facie case

If what we just said is correct, then in order for one to have a single representational capacity, one must have a whole 'battery' of them: representational capacities come, 'not singly, but in articulated networks.' [94] But it is very hard to believe that, for a creature to have a woozy feeling, or a tickle, or an itch, or a morphine-induced good feeling, it must be in possession of an articulated network of representational capacities.[95]

These considerations make it very hard to believe that intentionalism is true. As we noted earlier, if intentionalism is correct, then whatever phenomenal capacities are creature has, it has them wholly in virtue of having certain representational capacities. So if intentionalism is correct, then a creature has the ability to experience (say) a woozy feeling wholly in virtue of having certain representational capacities. Now we just argued that representational capacities come only in networks, not singly. So if intentionalism is correct, then a creature has the ability to experience a woozy feeling wholly in virtue of having a vast network of representational capacities. *So a creature's ability to experience a woozy feeling presupposes a vast network of representational capacities.* But the italicized statement is very hard to believe. It is very hard to believe that so much representational

horse-power is needed to have a woozy feeling or a feeling of nausea.

To be sure, there are some kinds of phenomenal content that a creature *must* have sophisticated representational capacities to enjoy. Consider the feeling a music lover has when listening to Chopin's ballade in G-minor, or the feeling and art-lover when looking at a Rembrandt. But intuitively it seems unlikely that one's ability to experience *any* kind of phenomenal content presupposes a whole slew of representational capacities. If we are right about this, then one's ability to experience certain kinds of phenomenal content is *independent* of one's ability to have mental entities with any representational content. And this would imply that one could have mental entities that had phenomenal content but *no representational content*; which means that phenomenal content is something altogether different from representational content – not identical with it, or supervenient on it.

I should address an objection:

You say that a creature whose mind is empty of representational capacities could experience an itch or a tickle. Well, an itch or a tickle must be experienced *as* being in a particular body part; so to experience an itch or a tickle, one must have the ability to represent different body parts to oneself. [96] Surely this ability *does* involve having a network of representational capacities. So, contrary to what you see, the ability to experience a phenomenal content *does* involve being in possession of such a network. [97]

There are a couple of points to make in response to this: a small one and a big one. Here is the small one. Not all phenomenal contents have felt bodily-location. Consider a morphinated feeling of euphoria. Is that felt in one's

hand or foot? No; it is a 'generalized' feeling. So it doesn't seem that any concepts of body-parts are involved in this kind of feeling. In fact, no spatial representations -- or *any* representational capacities, for that matter – seem to be involved the having of such a sensation.

The big point is this. To be sure, many sensations are felt *in* certain body parts, and their possession therefore *does* involve some ability to represent spatial location. But even when we confine our attention to sensations of that type, the facts just don't square with the supposition that phenomenal content supervenes on representational content. More specifically, we will find that if intentionalism were true, then sensations – even those that must be felt *in* certain body parts – would deliver vastly more information than they in fact deliver.

So, for example, we will see that, if intentionalism were true, then a tickle felt in one's knee would deliver far more information to the subject than such a sensation does in fact deliver; it would tell the subject far more than that there was some kind of disturbance in his knee – and, in fact, it would transmit far more information to the subject than it would be even remotely plausible to see it as transmitting.

In a word, intentionalism over-predicts the amount of information borne by sensations, even those that involve reference to body parts. Let us refer to this thesis as (OP).

Our argument for OP will consist of three parts. First we will consider the hypothesis that the phenomenal content of sensations – even those that involve reference to specific bodily locations – supervenes on *propositional* content. We will find that this hypothesis simply isn't consistent with the amount of information *actually* borne by such sensations.

Next we will consider the hypothesis that the phenomenal content of sensations supervenes on *non*-propositional representational content. And we

will find that *this* hypothesis simply isn't consistent with the amount of information actually borne by sensations. For argument's sake, we will provisionally take it for granted that there is such a thing as non-propositional content.

Finally, in the last section of the paper, we will see some reason to believe that *all* representational content is essentially propositional.

So our argument for OP will have the following form. Phenomenal content doesn't supervene on propositional content. Phenomenal content doesn't supervene on *non*-propositional, *supposing that there is such a thing.* And, finally, there is some reason to deny that there is such a thing as non-propositional content – so our argument t that phenomenal content doesn't supervene on propositional content is, if cogent, enough to prove that phenomenal content doesn't supervene on representational content *at all*.

Why phenomenal content doesn't supervene on propositional content

Consider some one part of your body; say, a specific part of your knee. Supposing that you are anatomically normal, you are capable of having an extremely wide range of sensations in that one part of your knee. Indeed, given any one *type* of sensation – say, the type *twinge* – you are capable of having a wide range of sensations of that one type, in that one part of your knee. As a squash player, I've injured my knees many times; and given practically any part of my knee, I've had many different twinges there. I don't think that any two of those twinges have had *quite* the same phenomenal content. Now remember that, according to the intentionalist, there is no difference phenomenal content of one's experience without there being some concomitant difference in the representational content of one's experience. So, according to the intentionalist, any two of the twinges I had

in my knee supervened on graspings of *different* propositions. In general, according to the intentionalist, any two qualitatively different sensations – no matter how minute the difference in question – correspond to graspings of distinct propositions.

This means that, if indeed the intentionalist is correct – if indeed the having of sensations (twinges, tickles, etc.) corresponds to the grasping of propositions – it follows that, typically, the propositions grasped will be highly *specific* ones; they will *not* be highly generic propositions like *there is damage to my knee.*

Does it make sense to suppose that the having of (e.g.) a twinge involves the grasping of highly specific propositions? There is reason to think not. What kind of proposition might be borne by a twinge in one's knee, given – what we've seen reason to believe – that this proposition must be fairly specific (cannot be anything nearly as generic as *my knee has been damaged in some way or other*)? Presumably it is some proposition about the condition of one's knee: about certain tissues, whether/how they are misaligned/inflamed/torn/…To put this another way: if a twinge in one's knee bears *any* proposition, it can only be assumed to be a proposition *about* one's knee, about its condition. So given that this proposition must be more specific than *there is some kind of damage in such and such part of my knee*, then presumably that proposition must, to some degree, *specify* the *kind* of damage or -- if the condition in question is not one of being damaged -- the *kind* of state one's knee is in. And, presumably, there is no way this information can be specified except in terms of the components of one's knee – the tissues, ligatures, bone, and so on – that make it up. So the proposition borne by the twinge will involve reference to quite specific states born by quite specific structures.

So, if intentionalism is correct, then *merely by virtue of having a twinge in*

one's knee, one can acquire a quite specific idea of what tendons, cartilage, and so on are, and a quite specific idea of the kinds of states that might befall them. But it is plain that a person could have a twinge and have absolutely no idea what e.g. a tendon was.

(Of course, to someone who had some *independent access* to knowledge of anatomy, a twinge in the knee might indicate some specific kind of damage. So e.g. a doctor who was an expert on knees might learn from a certain twinge that he had a highly specific kind of damage to his knee. But here ulterior or collateral information – information not derived solely from the twinge – would be at work. Strictly speaking, the twinge *by itself* wouldn't bear a proposition about a *specific* kind of damage; that would be done, not by the twinge itself, but by the doctor's *correlating* that twinge with independently obtained knowledge.)

Also, it simply *seems* wrong to suppose that e.g. mere tickles and twinges bear such heavy informational loads. A tickle seems to be much too simple and homogenous a thing to bear a fine-grained proposition about muscle exhaustion and misaligned tendons. Propositions are *complex*. To grasp a proposition, one must grasp its various constituents. (To grasp the proposition *Bob hit Bill* one must presumably grasp Bob, Bill, and the relation of hitting. Further, one must arrange these constituents in a certain way.) In any case, this seems like a not unreasonable proposal. Now could something as homogenous and amorphous as a twinge constitute a grasping of some highly structured proposition about sometimes quite recondite anatomical structures? Intuitively, it very much seems not.

(The intentionalist might say that, although twinges, tickles etc. *appear* too simple to mediate complex propositions, that appearance is misleading: how mental entities seem to be is not always how they are, *pace* Descartes. I will address this objection in a moment.)

To sum up, it isn't reasonable to suppose that the having of sensations supervenes on/reduces to the grasping of propositions. The propositions in question would have to be highly specific ones – ones loaded with differential information – and there is every reason to think that this is not the case.

I should respond to one last objection before ending this long section:

You have said repeatedly that bits of phenomenal simply don't *seem* to have the internal complexity needed to bear complex propositions -- propositions or factive propositions containing anything other than highly generic information. Presumably, you are basing this comment on your own reflection, on the 'data of introspection'. But such data are notoriously unreliable. Haven't we learned from cognitive science, and perhaps also from psychoanalysis, how little we can know of our own minds through introspection?

This point is well-taken. But there are a couple of points to be made in response. First of all, I should say that *I agree* that vast portions of our mental landscapes cannot be known through introspection. But surely if we can know *anything* on the basis of introspection, it is the nature of our phenomenal contents. If introspection is not completely useless as a means of gathering information about ourselves, it is probably pretty good at apprising us of the nature and character of our phenomenal states. And, given what introspection tells us about our phenomenal contents, it is hard to believe that they can bear the complex propositions that, if intentionalism is correct, they would have to bear. To my mind, there is nothing *intuitively* strange about mental *computational processes* (like those posited by Chomsky and Marr) that are completely unknowable through introspection. But the idea that I

might be similarly blind to e.g. the pain I'm feeling in my shin right now (having just banged it) is, to say the least, a radical one (though it has had its proponents). Of course, the objector could say that I am just begging the question: what is at issue is whether introspection is a valid source of information about our phenomenal contents, and I am simply assuming that they are.

But – this is the third counter-objection -- there is a more specific rejoinder to be made, and this one is not question-begging. If intentionalism is right, then phenomenal contents *deliver information.* This means that our minds must have some way of *retrieving* this information from them. So our minds must, at some level, be aware of the complexities, the articulations, that (if intentionalism is correct) enable them to bear this information. Could this awareness be *sub-conscious?* It seems not. The notion of phenomenal content is *very* closely bound up with that of consciousness. It seems that if phenomenal contents deliver information to *any* part of our minds, then presumably it is to our conscious minds. So *if* phenomenal contents deliver propositions to us, it is presumably the conscious mind that retrieves these propositions; and in this case, presumably, the conscious mind must have a pretty good grip on the character and structure of our phenomenal contents. So if intentionalism is right, then it seems to follow that our conscious minds *do* know our phenomenal contents pretty well. So given that consciousness tells us that our phenomenal contents lack the structure needed to mediate the heavy-duty propositions that, if intentionalism is right, they do in fact mediate – given this, we *ipso facto* have strong evidence that intentionalism is false.

Why the phenomenal content of sensations doesn't supervene on non-propositional content (supposing that there is such a thing)

So far we've seen some reason think that phenomenal content doesn't supervene on the grasping of *propositions.* Some philosophers hold that the kind of representational content embodied in propositions is not the only kind of representational content: they say that there is also 'non-propositional' representational content. (For some reason, the term 'non-conceptual' is typically used instead of 'non-propositional.' But I will use the term 'non-propositional.'[98]) Obviously *if* there is such a thing as 'non-propositional' representational content, then we have not shown that intentionalism *tout court* is false, only that a specific kind of intentionalism is false: for the intentionalist could counter-attack by saying:

Phenomenal content supervenes on *non*-propositional content. Whenever there is a difference in phenomenal content, that fact *does* correspond to a difference in representational content. It corresponds to a difference in *non*-propositional content. So, for example, if two tickles are qualitatively a bit different from each other, that means that they supervene on states that bear different *non*-propositional content. Thus, intentionalism *is* true, after all: whenever there is a difference in phenomenal content, there is a difference in representational content. But the latter is *non*-propositional, not propositional. [99]

In this section, I will show that *even if there is such a thing as non-*

propositional content, that doesn't threaten our defence of OP. So *if* there is such a thing as non-propositional content, that does not (so I will argue) enable the intentionalist to sidestep the arguments we've already set forth.

(In the next section, we will question the hypothesis that there is such a thing as non-propositional content.)

The main things we said about propositional content are true of non-propositional content (supposing for the sake of argument that there is such a thing as non-propositional content). And the arguments we used earlier – when we were attacking the idea that phenomenal content might supervene on *propositional* content – work with equal force against the idea that it might supervene on non-propositional content. Accordingly, what we say here will closely parallel what we said earlier.

Consider any given perception – e.g. the visual perception you are having right now of the journal in front of you. The journal is being given to you *as* having innumerable properties – *as* having a certain colour, shape, size, spatial location, and so on.

For the sake of argument, let us suppose that there is such a thing as non-propositional content and that, consequently, some representations bear non-propositional content. Surely what we said a moment ago is true of representations that bear non-propositional content (supposing that there are such things). Surely such a representation must represent things *as* being a certain way and thus *as* having certain properties. It is a contradiction in terms to say that there is a representation that *fails* to represents things *as* being a certain way. So *if* there are representations that bear *non*-propositional content, *those* representations must – like any other kind of representation – represent things *as* being a certain way and thus *as* having such and such properties. So the ability to have *any* kind of representation -- regardless of whether those representations bear propositional or non-

propositional content – presupposes the ability to represent things *as* having properties; it presupposes possession of what we earlier referred to as 'representational capacities.'

Earlier we argued that, for any property P, if a creature can represent things as having P to itself – if it can see things/ hear things/ imagine things/ …*as* having P – then it necessarily has the ability to represent things to itself as having various other properties. Representational capacities, we argued, come not singly, but in vast, interconnected networks. The argument was basically that a creature cannot represent things as having P to itself unless it can 'locate' P in a 'space' of related, but distinct properties. Now as we just noted, *any* representation – whether it bears propositional or non-propositional content – presupposes the ability to represents things *as* having certain properties to oneself. So one cannot have mental states that bear non-propositional content (supposing that there is such a thing) unless one possesses a "vast, interconnected" network of representational capacities.

Given this, suppose that, in fact, phenomenal content *does* supervene on non-propositional content. In that case, given what we just said, the ability to have a tickle, or a twinge, or a feeling of nausea presupposes the ability to represent things *as* having various properties to oneself, and therefore presupposes that one is in possession of a vast, interconnected network of representational capacities. So, to reiterate, *if* phenomenal content *does* supervene on non-propositional content, then the ability to experience any quale – any twinge, or tickle, or woozy feeling – presupposes that one is in possession of a vast arsenal of representational capacities.

But this is very unlikely to be the case. Intuitively, it is very hard to believe that the having of a feeling of nausea, or of a morphinated high presupposes possession of a vast network of representational capacities.

Of course, the intentionalist could say: 'You are just begging the question.

Given that representational capacities come in vast networks, what follows is that the ability to have (e.g.) a tickle or a twinge *does* presuppose possession of a veritable battery of representational concepts.'

But there is good reason to be suspicious of this counter-move. By definition, intentionalism says that *each* difference in phenomenal content corresponded to a difference in representational content. Now as we noted earlier, given how many different phenomenal contents we can have – indeed, given how many different phenomenal contents one can have *in practically any given part of one's body* -- it follows that if *each* difference in phenomenal content corresponded to a difference in representational content, then each bit of phenomenal content – each twinge, each nauseas feeling, each tickle – would have to contain an enormous amount of differential information and would thus be enormously informative. If intentionalism is correct, then any difference between two phenomenal states – even the merest nuance of a difference -- corresponds to a *representational* difference. Now our phenomenal 'palate' is extremely rich: we seem capable of having indefinitely many different kinds of phenomenal contents in practically any given body part. In fact, we seem to be capable of having indefinitely many phenomenal contents *of a given kind* in a given body part. There are indefinitely many different kinds of tickles, or twinges, or pains, or itches that one can have on (say) a given part of one's neck. For intentionalism, each of these phenomenal differences corresponds to an informational difference. So given how many different kinds of phenomenal contents we can have – given how many different kinds of such content we can have in practically any given body part – any phenomenal content would have to contain an enormous amount of differential information, and would that have to be enormously information-heavy. Given how information-heavy a tickle or twinge would have to be, for intentionalism to be correct, it follows that a

tickle or twinge would *not* bear highly generic messages like *there is a disturbance of some kind in such and such part of my body*, but would bear much more fine-grained, precise messages.

But a tickle or a twinge simply seems *not* to bear that much information. Here we need do little more than repeat what we said earlier. For the sake of argument, suppose that, in fact, a tickle *did* contain such a large amount of information. What would that information be *about*? If (e.g.) a tickle were in fact a treasure-chest of differential information, what would that information be information *about*? It could only be information about some bodily condition– information about some somatic disturbance or event. So *if* intentionalism were correct, then (e.g.) a tickle in one's neck would tell one an enormous amount about the condition of that part of one's neck: it would do much *more* than to give one some highly generic message like *disturbance of some kind in such and such a part of my neck.* But tickles and twinges simply don't tell one that much.

We said all of this before. The important point is that the argument we just gave goes through regardless of whether we are talking about propositional or non-propositional content. Even if the kind of information borne by (e.g.) a tickle is *non*-propositional, it *still* follows, for the reasons just given, that intentionalism would require of that tickle that it bear an enormous amount of differential information. But tickles simply *don't* bear that much information. They don't bear that propositional information – as we saw before – and they don't bear that much *non*-propositional information. For even if (say) tickles and twinges bore a large amount of *non*-propositional information, that information would still, inevitably, be about one's body – about somatic disturbances: so each tickle or twinge would constitute an extremely high-resolution, information-heavy apprehension of the condition of a given part of one's body. But while it *could* credibly be maintained that tickles and

twinges give one *some* information about one's body, it cannot credibly be maintained that they constitute high-resolution apprehensions of bodily conditions: in so far as tickles, and twinges, and feelings of nausea are representations, they are fairly opaque and uninformative. Intentionalism – whether it construes phenomenal content as supervening on propositional or *non*-propositional information – demands that tickles, twinges, etc. be very much non-opaque and high-resolution. So intentionalism, of either of the two varieties just mentioned, simply lacks credibility.

A reason to doubt that there is such a thing as non-propositional content

So far we have allowed, for the sake of argument, that there is such a thing as non-propositional content. Here I will argue, in effect, that propositional content is the *only* kind of representational content. If this is the case, then of course it cannot be maintained that phenomenal content supervenes on non-propositional content. Consequently, our earlier argument that phenomenal content cannot supervene on *propositional* content would, if cogent, amount to a demonstration that phenomenal content cannot supervene on *any* kind of representational content.

I warn this section is *highly* speculative and is in no way probative. Its purpose is not to *refute* the idea that there is non-propositional content, but only to put some pressure on this idea – to delineate some of the obstacles that view will run into.

Once again, consider an arbitrary visual perception. (Supposing that, indeed, some representations have non-propositional content, visual perceptions are paradigm cases of such representations; so we might as well focus on them.) Uncontroversially, the representational content of a visual

perception can *to some degree* be put into words and, thus, embodied in a proposition. People can tell you what they saw, at least up to a point ('I saw a man in an orange shirt smoking a pipe…') So obviously the representational content of a visual perception overlaps *to some extent* with the representational content of a sentence ('I saw a man in orange shirt…'): a paradigmatically 'digital' representation. Again, I do not think this is controversial: who could possibly maintain that the representational content of a visual perception is utterly ineffable – is totally incapable of being embodied in a sentence? Surely when I say 'I saw a red apple sitting on a desk…' I am succeeding in transmitting *some* of the representational content of a visual perception of mine.

Now the next step is also, I think, fairly uncontroversial. Visual representations are, in an important sense, homogeneous. When we consider a visual perception, there is no reason to believe that it has two, totally different kinds of representational content; visual perceptions seem to be (in this respect) 'seamless'; the representational content of a visual perception seems not to be patchwork of totally different kinds of content – of propositional and non-propositional content. Visual perceptions are not cobbled together out of discrete parts.

What I mean by this can best be understood by contrasting visual representations with patently hybrid –- patently heterogeneous --- representations. A comic strip is a patently hybrid kind of representation: there is a picture (e.g. of a superhero performing some great feat) and there are also words (which are put in a bubble emanating from the mouth of the featured character). So comics are both analogue and digital and are thus, in an obvious way, heterogeneous representations; they are characterized by a certain kind of internal division. Now when we consider *purely* analogue representations – e.g. visual perceptions, photographs – there seems to no

comparable internal division. Analogue representations have a kind of seamlessness, a kind of homogeneity. And, one would think, a representation would not have this seamless, homogenous quality if it were built up out of two, totally different kinds of content.

In any case, there is a great deal of intuitive support for the idea that a visual perception or a photograph or a painting does *not* consist of two different kinds of representational content – for they do not have that 'cobbled together' quality characteristic of e.g. comics. There is simply no intuitive support for the idea that visual perceptions and photographs bear two, totally different kinds of content.

Hence, the facts we have to work with are as follows. (i) The content of an analogue (e.g. a visual) representation can, at least up to a point, be expressed by a sentence and thus, presumably, embodied in a proposition. (ii) An analogue representation does not bear two, completely distinct kinds of representational content.

Now given both (i) and (ii), it seems to follow that the representational content of a visual perception is *in its entirety* propositional. But this conclusion is in deep tension with the following fact, viz.: (iii) There is, plainly, *some* kind of important distinction between analogue (e.g. visual) and digital representations (e.g. propositions): the latter *are*, whereas the former *are not*, decomposable into minimal representational units.

The best way to make sense of these three facts is, I would suggest, as follows: the difference between propositions and e.g. visual representations is not a difference in *what* is represented but in *how* it is represented. Each of (i)-(iii) is correct. The difference between e.g. a sentence and a visual perception lies not in what is represented, but in how what is represented is *packaged.* A sentence tidies everything up, putting it in discrete parcels. A visual representation leaves it, as it were, strewn all over the floor. A

proposition is not a special kind of information; it is a special *packaging* of information.

I think that this line of thought does less violence to our intuitions than the alternatives. First of all, we *do* want to say that the information borne by a visual perception can (at least up to a point) be embodied in a proposition. But we *don't* want to say that the representational content of a visual perception is a proposition (or conjunction of propositions), for the simple reason that propositions are, whereas the contents of visual representations are not, neatly decomposable into minimal units. (Surely the proposition *Bob is smarter than Fred* is decomposable into discrete parts while the representational content of a visual perception of a forest is not thus decomposable.) We also probably don't want to say that a visual perception has two wholly distinct kinds of representational content: propositional and other. Given these data, it seems reasonable to say the following: A visual (more generally, an analogue) representation bears only kind of representational content; and this content *can* be packaged in a proposition, but it doesn't have to be.

So it is true that sentences bear propositions while visual perceptions do not; but it is also true that the content of a visual perception can be embodied in propositions. If propositions were *one* kind of representational content while the representational contents of visual perceptions were some totally different kind of content, this would not be possible. But propositions are not a kind of content; they are a way of packaging content. Visual representations don't package content in this way; but what they do package can be packaged propositionally.

Now if this is even remotely on the mark, then there are two quite distinct dimensions to the 'representational load' of a representation: there is the representational content of that representation; and there is also the *packaging*

of that content. Now the packaging of this content has too often been identified with the representation itself. But this is a mistake. A visual perception of a round ball and the sentence ‘that ball is round’ obviously overlap in what they represent; but the way they package this common content differs. So what is borne by a sentence is not *just* representational content, but a certain way of packaging this content; and the same is true of a visual perception. (The sentence itself is not identical with the packaging. The sentence ‘Bob is intelligent’ packages the same representational content in the same way as the sentence ‘Bob is smart’. But these sentences are obviously different.) But what it is that is packaged is not a proposition: and that is why visual representations – and, more generally, analogue representations: representations which do not have propositions for the representations contents – can have their contents embodied in propositions.

With these points in place, we can neutralize the objector’s point. Earlier in the paper we showed that phenomenal content probably doesn’t supervene on the kind of representational content embodied in *propositions.* Now we’ve just seen that, in all probability, there are not *two* kinds of representational contents: there is only one.

(There *is* an important distinction between e.g. visual perceptions, on the one hand, and sentences, on the other: but this difference has to do *not* with the representational content borne by such things, but by the *way* in which this representational content is packaged. The difference that is *referred to* as ‘the difference between conceptual and non-conceptual content’ *does* exist, but it is not a difference between kinds of content, but between ways of packaging that content.)

So since (i) there is *not* a difference in the kind of representational content borne by propositions, on the one hand, and visual perceptions (and photographs) on the other (though there is a difference in the way these

things ‘package’ their representational content); and (ii) since, as we’ve seen, the having of phenomenal content probably doesn’t supervene on the grasping of the kind of representational content embodied in propositions; it follows by simple logic (iii) that the having of phenomenal content doesn’t supervene on the kind of content embodied in visual sensations, photographs. And once (iii) is granted, the intentionalist counter-attack put forth by the objector (page 23) ceases to have any viability.

There is a less formal, more intuitive, way of nullifying the objector’s counter-attack (p.23). Suppose for the sake of argument that *there is* such a thing as non-propositional content. Now *if* there is such a thing as non-propositional content, then indeed phenomenally pregnant mental states – e.g. sense-perceptions, sensations (twinges, tickles, etc.) – are presumably the kinds of things that *would* bear such content. Given this, consider any given phenomenally pregnant state – consider e.g. a tickle. Now a tickle does, it would seem, tell you *something.* It tells you that there is *some* kind bodily disturbance in a certain region of your body. Now consider the information borne by the tickle, and subtract away all the information that can be embodied in *propositions.* The first thing you’d have to subtract would be a reference to any bodily location. For obviously information about bodily location can be embodied in a proposition. In fact, you’d have to subtract *any* kind of spatial or even temporal information: for all such information can be embodied in propositions. So any information, in that sensation, to the effect that such and such is happening *at such and such a time* and *at such and such place* – all that information would have to be removed. For these things can be stated in words, and thus embodied in propositions. Of course, once information about spatial and temporal location is gone, it isn’t really clear what kind of information is left over: for, it would seem, *any* kind of information about the external world will *presuppose* some kind of spatio-

temporal information. So once *propositional information* is removed from a phenomenal content – once the kind of information that can be embodied in propositions is taken away – there doesn't really seem to be *any* information left.

Our 'framework' concepts – the concepts in terms of which the external world is given to us: space, time, object, event, change etc. – are all capable of being embodied in propositions. For nothing is more obvious than that we talk about space, time, change etc. So given a mental state, once we remove the 'propositional' information from it, we *ipso facto* divest that state of its being able to tell us *anything* about the world. This consideration obviously favours the thesis that there is no non-propositional content.

Another example might help. Consider a visual perception. As we've noted, it is undeniable that *some* of the information borne by that perception can be embodied in propositions. I can tell you what I see, at least up to a point: I see 'a man wearing a green shirt, holding a tennis racket' etc. Now strip away all the information from that perception that can be embodied in propositions. What are you left with? Not with a perception of an object of such and such *shape*, occupying such and such a *place.* For this information is clearly of the kind that can be embodied in propositions: I can obviously *tell* you that I see a man of such and such shape in such and such a location etc. But once *this* information is taken away, it isn't really clear if anything representational is left. Once all propositional information is taken away, the possibility of *any* kind of representation is taken away. Once I remove from the visual perception all reference to things that can be stated – which includes location and state of motion – it very much seems that the pre-condition for there be *any* kind of representation is taken away.

In sum, the kind of information that can be embodied in propositions must be borne by *any* representation worthy of the name. There is no way to

subtract propositional information from a representation without making it cease to be a representation: in effect, you can't get rid of propositional content without getting rid of *all* content. This, of course, favours the conclusion that there is *no* non-propositional content.

Now one might reject the last inference:

Given only that you can't get rid of propositional content without getting of *all* representational content, it doesn't follow that all propositional content is propositional. This inference involves a non-sequitur. You can't get rid of extension without getting rid of colour. But extension and colour are different things.

There are two points to make in response. First of all, I grant that technically the objector is right. Given *only* that propositional content cannot be removed from a representation without divesting that representation of all content – given *only* that fact, it doesn't *technically* follow that all representational content is propositional content. But surely that fact gives *some* credence to the idea that all content is propositional content.

But we have a stronger counter-response at our disposal. *Even if it is granted that there is such a thing as non-propositional content,* once it is conceded that a mental entity cannot be stripped of its propositional content without being stripped of *all* its representational content, the idea that phenomenal content supervenes on non-propositional content becomes null and void. Permit me to defend this thesis.

We've just seen that non-propositional content, if such a thing exists, is not autonomous, but is necessarily mixed together with propositional content. This seems to imply that two mental states cannot differ in *non*-propositional content (supposing for the sake of argument that there is such a thing)

without also varying in propositional content. After all, if two mental states *could* vary in non-propositional content without varying in propositional content, that would very much seem to indicate that non-propositional content has just the kind of independence form propositional content that we have seen that it could not possibly have. So if there is such a thing as non-propositional content, then given that it is parasitic on propositional content, it follows that two states could not differ in non-propositional content without also differing in propositional content. So any difference in non-propositional content is associated with a difference in propositional content. Now the version of intentionalism we are considering now says that differences in phenomenal content correspond to differences in *non*-propositional content. We've just seen that differences in non-propositional content correspond to differences in propositional content. So, by implication, the kind of intentionalism we are considering now implies that differences in phenomenal content correspond to differences in *propositional* content. But we've seen, over and over again (pp.11-22), that this is not the case.

In sum, *even if* there is such a thing as non-propositional content, this doesn't help the intentionalist: for a change in non-propositional content (supposing that such a thing exists) involves a change in propositional content; consequently, if differences in phenomenal content correspond to differences in *non-propositional* content, then they also correspond to differences in *propositional content.* But we've seen many reasons to reject the idea that differences in phenomenal content always correspond to differences in propositional content.

References

Brentano, Franz

[1] *Psychology From an Empirical Standpoint* edited by Linda L. McAlister with a new introduction by Peter Simons. Translated by D. B. Terrell, Antos C. Rancurello, and Linda L. McAlister. London and New York: Routledge, 1995.

Byrne, Alex

[1] 'Intentionalism Defended'

The Philosophical Review, Vol. 110, No. 2 April 2001

Dretske, Fred

[1] *Naturalizing the Mind* M.I.T. 1995

Evans, Gareth

[1] *The Varieties of Reference* edited by John MacDowell Oxford 1984

Fodor, Jerry

[1] *Concepts* Clarendon Press Oxford 1998

Harman, Gilbert

[1] 'The Intrinsic Qualities of Experience' in *Philosophical Perspectives 4*, 1990.

Heck, Richard G.

[1] “Nonconceptual Content and the ‘Space of Reasons’” *The Philosophical Review*, vol. 109, No. 4 (October 2000)

Heidegger, Martin

[1] *Being and Time*, transl. by J. Stambaugh, SUNY Press 1996

Lycan, William

[1] *Consciousness and Experience*. M.I. T. Press 1995

MacDowell, John

[1] *Mind and World* Harvard 1994

Nagel, Thomas

[1] ‘What is it like to be a bat?’ in *The Nature of Mind* edited by David Rosenthal Oxford 1990

Peacocke, Christopher

[1] *A Study of Concepts* MIT 1992

[2] *Sense and Content* Oxford 1983

[3] 'Holism' in *A Companion to the Philosophy of Language* edited by Bob Hale and Crispin Wright Blackwell 1997

Sartre, Jean-Paul

[1] *Being and Nothingness: A Phenomenological Essay on Ontology*, tr. by Hazel E. Barnes (Washington Square, 1993)

Sellars, Wilfrid

[1] '*Empiricism and the Philosophy of Mind* ' originally published in Minnesota Studies in the Philosophy of Science, vol. 1 1956; reprinted in *Science, Perception and Reality* Routledge and Kegan Paul London 1963

Tye, Michael

[1] *Ten Problems of Consciousness* MIT 1995[JK9]

[2] *Consciousness, Color, and Content* MIT 2000

Simulated versus actual thought

Part I. In this paper, I wish to discuss the relation between *simulating* thought, on the one hand, and *actually* thinking, on the other. I believe that this relation has been distorted by a failure to take into account some basic facts about the conditions under which separate thoughts can combine to form a single molecular thought.

Let us start with a story. At one moment, you think *snow is white.* At a later time, you think *something is white*. But your thinking *something is white* has nothing to do with your previously thinking *snow is white.* Under these circumstances, have you thought *snow is white, therefore something is white*? Have you thought *snow is white and something is white*? Have you thought *any* molecular proposition comprising both *snow is white* and *something is white*?

"No" to all of these questions. Because your thinking *snow is white* didn't have anything to do with your subsequently thinking *something is white*, your two thoughts didn't "combine", so to speak, into a single thought-process.

Let us extend these remarks by telling another story. You think *Socrates*. Later you think *hit*. Still later you think *Plato.*[100] But your thinking *Socrates* has nothing to do with your thinking *hit*, and your thinking *hit* has nothing to do with your subsequently thinking *Plato.* Under these circumstances, have you thought *Socrates hit Plato?* You have not. In order to think *Socrates hit Plato*, it is necessary not only to have the sequence of thoughts *Socrates*hit*Plato* (each * is meant to indicate a break of a kind): it is necessary also that your thinking *Socrates* have something to do with your thinking *hit* and also with your thinking *Plato.*

There is a difference between having a series of cognitive episodes, on the one hand, and having a *unified* thought-process, on the other. Not just any

sequence of representations constitutes a thought-process. One's thinking such and such must have something to do with one's later thinking thus and such, if one is to think *such and such, therefore thus and such* or *such and such and thus and such* or *if such and such, then such and such*, and so on.

These banal facts give us some leverage on a question that is anything but banal, namely: "Is there a difference between *simulated* thought and actual thought?"

Let me break the question down. Let M be some physical object in which there are representations or symbols of some kind. We may leave it open whether M is an artifact, like a computer, or a naturally occurring object, like a brain. Because of M's structure, the representations it produces follow one another in utterly "rational" ways – in ways that are consistent with the strictures of logic. If M produces a *snow is white*-representation, then – as an inevitable and perfectly reliable consequence of M's structure – M will duly produce a *something is white*-representation. If M produces a *2+2*-representation, then M will duly produce an *=4*-representation. In general, for any chunk of information *i*, if M produces an *i*-representation, then M inevitably produces an *i**-representation, where *i** is a chunk of information that stands in an appropriate logical relation to *i*.

Here is the question I am asking: "Given only that M reliably follows up representations with other, logically appropriate representations, does it follow that M is rational?" My answer is: "No." Given only that M perfectly *simulates* rational thought, it doesn't follow that M is rational. Of course, M might still be a rational thinker for some other reason.[101]

Part II. Let us begin with some pedestrian facts about causality. Causal relations hold among states of affairs, not among things. The *rock* doesn't

break the window. Some state of affairs involving the rock does so. The rock has a certain mass; it is moving with a certain velocity; it collides with a window of a certain thickness. *These things* – the rock's having a certain mass, moving with a certain velocity etc. – are what cause the window to break. The rock *simpliciter* does not do so.

This is equivalent to saying: it is not the rock *per se* that breaks the window. It is the rock's *having certain properties* that does so. It is the rock's moving with a certain velocity, colliding with a window of a certain thickness, and so on.

Not every property of the rock is involved in the window-breaking. Suppose the rock has a dimple on one of its sides. The rock's having that dimple *is* causally effective in some contexts. For example, if you were to pour water on the dimpled side of the rock, the presence of the dimple would affect the course of the stream of water. But the dimple needn't be causally effective in this particular context. Perhaps the side of the rock that collided with the window wasn't the side with the dimple on it.

The main point is that an *object* doesn't cause anything. What does so is an object's having this or that property; equivalently, it is a state of affairs comprising that object.

Let us apply these platitudes about causation to the nature of thought-processes. It might seem that any sequence of mental representations constitutes a thought-process. First there is the thought: *snow is white.* Then there is the thought: *grass is green.* Then there is the thought: *snow is white and grass is green.*

Here we have a sequence of mental representations. That, it might seem, is enough for the existence of a thought-process. Indeed, in this particular case, we have enough for the existence of a *rational* thought process – a proof or logical derivation. A thought that *snow is white and grass is green* occurred

after a thought that *snow is white* and a thought that *grass is green.* So that conjunction was appropriately derived from those conjuncts.

This reasoning is spurious. There was not necessarily any thought-*process* at all, let alone a rational one. Consider the following scenario. X thinks *snow is white* at t. (Let C be this thought, i.e. this cognitive episode.[102]) Y thinks *something is white* at t*. (Let C* be this cognitive episode.) Y's thought has nothing to do with X's. The two thoughts happen independently. There is no causal connection. Has a single thought-process occurred? Has there been a thought to the effect that *snow is white, something is white*? No. There have been two disjoint thoughts. The two thoughts didn't combine into a single thought-*process*. The occurrence of the second thought didn't have anything to do with the occurrence of the first. The lack of a causal connection prevents that.

It seems, then, for thoughts to "molecularize", a causal connection between them is necessary. But not just *any* causal connection will do. We could easily imagine a case where Bill's thinking *grass is green* causes Fred to think *something is green*. It suddenly occurs to Fred that grass is green. For some random psychoanalytic reason, this causes Fred to dump green paint on Fred's car. Fred sees his car, and thinks (among other things) *something is green.* Here the sequence of thoughts *grass is green…something is green* doesn't constitute a thought-process, even though the first thought caused the second one to occur. So if distinct thoughts are to form a single thought-process, a causal connection is necessary, but not sufficient.

For illustrative reasons, we spoken about thoughts had by different people. But what we've said so far can apply to thoughts occurring within a single mind. To see this, let us tell a new story. For whatever reason, Bill thinks: *grass is green.* A moment later, out of the blue – for reasons totally unrelated to his having the first thought -- he thinks: *something is green.*

Here, within a single mind, we have a sequence of thoughts: *grass is green, something is green*. But that sequence of thoughts does not constitute a single thought-process, let alone a coherent one. The second thought has nothing to do with the first. There is no causal relation. That, among other things perhaps, prevents the two thoughts "molecularizing" into a single thought-process.

The propositions involved – *grass is green, something is green* – stand in the right logical relations to each other. So the reason Bill isn't having a coherent thought-process has nothing to do with a lack of bearing-relations among the propositions he believes. As we said, it has to do with (inter alia) a lack of a causal connection among the mental states embodying his beliefs.

Let us now change the story. Again, suppose Bill suddenly thinks *grass is green.* Let B_1 be that cognitive episode. A little later, Bill suddenly thinks *something is green.* Let B_2 be *that* cognitive episode. Of course, B_1 and B_2 are spatio-temporal entities: they are mental states, and they are also, presumably, brain-states.

For reasons we've seen, if the sequence B_1-B_2 is to constitute a thought-process, it is *necessary* that they be causally related. But it is not sufficient. Here we have to start taking into account the points made earlier about causation.

B_1 and B_2 are physical entities. As such, they have causal powers. So B_1 causes things to happen. But not every case where B_1 causes something to happen is a case where B_1's *being a thought that grass is green* causes something to happen. Because of its chemical properties, B_1 may cause you to have a headache or to feel elated. But those are not, at least not necessarily, cases where B_1's *being a belief that grass is green* causes anything to happen. Of course, B_1 *is* a belief that grass is green; and presumably that fact about B_1 *is* causally effective, at least in some contexts. But that doesn't

mean that, every time B_1 does something, B_1's being a belief that grass is green is involved.

Suppose, if only for the sake of argument, that B_1's having a certain shape is not involved in B_1's realizing that content. (If you think that shape *is* essential to representational content, then simply substitute my talk of shape with talk of some other property that you think to be incidental to B_1's being a belief that grass is green.) Given this, suppose that B_1 causes B_2 to come about, but that B_1's *being a belief that grass is green* is not involved in this transaction. So what is causally effective is B_1's having a certain shape, not its having a certain representational content.

Here we have the sequence B_1-B_2, but the generation of that sequence has nothing to do with B1's being a belief that grass is green. What we have here is a sequence of thoughts, where the first thought causes the second to exist, and where the representational content of the first thought entails that of the second. But we *don't* have a case of valid reasoning.

In fact, we don't even have a thought-process, let alone a valid one. Suppose that Bob is the person whose mind-brain houses B_1 and B_2. In the scenario described, Bob's *believing that grass is green* is causally inert with respect to his subsequently believing that something is green. So Bob's coming to believe that something is green has nothing to do with Bob's previously believing that grass is green. In order for the sequence B_1-B_2 to constitute a thought-process – whether logical or illogical – it is necessary not just that B_1 cause B_2 to come about, but that B_1's having a certain representational content be implicated in that transaction. So under the circumstances, Bob has not had a single thought-process.[103]

A thought process is not just a series of a series of mental representations. It is a series of representations such that, for each entry in that series, that

entry's *having a certain representational content* is causally effective in generating later entries.

Even that condition, though clearly necessary, is probably not sufficient. We can imagine analogues of points made by Davidson (in connection with action-theory[104]) and Gettier (in connection with epistemology[105]) showing that b_i's being a belief that such and such is what brings about b_{i+1}, but where, nonetheless, the sequence b_i-b_{i+1} is not a thought-process.

Let us illustrate this with a new story. (Let B_1 and B_2 be defined as before.) For whatever reason, Bob finds it terribly exciting that grass is green. So whenever the thought *grass is green* flashes through his mind, his heart starts to race, his palms start to sweat, and so on. Prior to all this, a hypnotist programmed Bob to think *something is green* whenever his (Bob's) heart rate went above ninety beats per minute. One day, the thought *grass is green* flashes through Bob's mind -- B_1 occurs. For the reasons just described, B_1's being a belief that grass is green – not just B_1, but B_1's being a belief of that kind – causes Bob to think *something is green.*

Nonetheless, it seems that, even in this circumstance, the sequence B_1-B_2 does *not* constitute thought-process. B_1's having a certain representational content is indeed causally effective. But it is it not, evidently, effective in the right way. So for the sequence B_1-B_2 to constitute a thought-process, as opposed to a sequence of isolated thoughts, it is not *sufficient* that B_1's having such and such representational content be causally effective in bringing about B_2.

But it is *necessary.* It is not enough that B_1 cause B_2 to happen. It is also necessary that B_1's *being a belief that grass is green* be causally involved.

To consolidate these points, it will help if we introduce some

terminology. We speak of certain entities as being “truth-evaluable”. Such entities include propositions, beliefs, and sentences. Some entities can be evaluated, not for truth, but for validity or invalidity, logicality or illogicality. [106] A series of thoughts can involve false premises but, aside from that, be impeccable. Such a series could be described as valid, but not sound. Ordinarily, we use the terms “valid” and “sound”, and the like, to describe sequences of propositions, not cognitive episodes. But it will expedite discussion if we now those terms, or homonyms thereof, to describe sequences of cognitive-episodes. Given anything that can be evaluated for logicality, let us describe it as logic-evaluable.

Using this terminology, we may summarize our findings thus. For a sequence of brain-states $b_1...b_n$ to constitute a logic-evaluable thought process, it is necessary that, for each i, b_i’s *having such and such content* be what generates later entries in that series. (This is necessary but not, as we just saw, sufficient.) Indeed, this condition must be met if $b_1…b_n$ is to constitute *any* thought-process, and not just a series of disjoint entries.

Part III. Let us continue to talk about our old friend Bob. Let us suppose we do a bit of neuro-surgery on Bob. We “tweek” Bob’s brain so that, when B_1 occurs, B_2 must also occur. Through some clever neuro-engineering, we make it a physical necessity that, if B_1 happens, Bob’s brain will produce B_2.

But there is a qualification. We make sure that what makes B_2 happen is not B_1’s being a belief that grass is green. So even though B_1 necessarily, and therefore reliably, leads to B_2, nonetheless B_1’s *being a belief that grass is green* has nothing to do with B_2’s coming into existence.

Perhaps this qualifier requires clarification. Let p be some property of B_1

that is not involved B_1's being a belief that grass is green. But suppose that, in general, p is a causally effective property. So p isn't a property like *is a thing x such penguins either do or do not fly*. p is, perhaps, the property of having a specific temperature or a specific micro-morphology. So even though it is causally effective, p is not involved in B_1's being a belief that grass is green. [107] Our neuro-surgeons set things up in such a way that it is B_1's having p – and *not* B_1's being a belief that grass is green – that is causally responsible for B_2's coming into existence.

Make no mistake. The causal connection between B_1 and B_2 is *reliable.* Our engineers rigged Bob's brain in such a way that, given the laws of physics, B_2 absolutely *must* happen if B_1 happens (at least in so far as anything *must* happen, given the apparently probabilistic character of physical law). So the connection between B_1 and B_2 is as reliable as a connection can be.

At the same time, B_1's causing B_2 to occur has nothing to do with B_1's having a certain representational content. It has to do entirely with some property of B_1's that is irrelevant to B_1's representational content.

So under this circumstance, Bob's thinking *something is green* has nothing to do with his thinking *grass is green.* True – Bob has a brain-state (B_1) encoding a belief that grass is green. True – that brain-state brings about a brain-state (B_2) to the effect that something is green. True – the first brain-state brings about the second via a reliable mechanism. But Bob's thinking *grass is green* doesn't have anything to do with Bob's thinking *something is green.* So Bob has not thought *grass is green, therefore something is green.* Nor has Bob thought *grass is green and something is green.* Bob hasn't thought any "molecularization" of those propositions. The sequence B_1-B_2 does *not*, under this circumstance, constitute a thought-process at all. *A*

fortiori it does not constitute a piece of valid reasoning. Even though the propositions in question – *grass is green* and *something is green* --- are being thought in the right order, nonetheless the sequence B_1-B_2 does not constitute a valid *thought*-process. It doesn't constitute *any t*hought process, valid or invalid.

What we said about B_1-B_2 applies to *any* sequence of brain-states – or, indeed, to physical states of any kind -- and, thus, to *any* sequence of mental representations. Let P and Q be any two propositions such that P entails Q. With that one qualification, there is no restriction on what P and Q may be. Let B_P be a brain-state of Bob's (or Mary's or Fred's…) mediating a belief that P, and let B_Q be a brain-state of the same person that mediates a belief that Q. Finally, let p* be some causally effective property of B_P that is not relevant to B_P's being a belief that P. Through clever neuro-engineering, we can ensure that, if B_P happens, then B_Q must happen. And we can also make sure that B_P's *being a belief that P* is not involved in that causal transaction: we set things up so that it is B_P's having p* -- and not B_P's being a belief - - that causes B_Q to come about.

For reasons we've seen, the sequence B_P-B_Q will not constitute a valid, or even a logic-evaluable, thought-process. In fact, it won't even constitute *a* thought-process. It will just be a series of disjoint entries.

So far we've been talking about brain-states. But exactly the same principle applies if we were to talk about states of a punch-card automaton or a Texas Instruments calculator or any other physical entity.

This line of thought can be extended to the sub-propositional level. Let b_1, b_2, and b_3 be the brain-states mediating, respectively, your thinking *Socrates*, your thinking *hit*, and your thinking *Plato.* If b_1's being a *Socrates*-thought has nothing to do with b_2's coming into existence, or if b_2's being a *hit*-

thought has nothing to do with b_3's coming into existence, then the sequence b_1-b_2-b_3 doesn't add up to your thinking *Socrates hit Plato.* You have thought *Socrates* and then thought *hit* and then thought *Plato.* But you have not thought *Socrates hit Plato.* For these three thoughts to molecularize, it is necessary not only that *they* be causally interrelated; it is necessary that their *having* certain representational contents be what drives the sequence. [108]

Through clever neuro-engineering, we could "tweek" Bob's brain in such a way that, if b_1 occurs, then b_2 and b_3 *must* occur, in that order, where the "must" corresponds to the strongest kind of nomic necessity imaginable. But as we have just seen, b_1's causing b_2 and b_3 to happen will *not* add up to the thought *Socrates hit Plato*, unless b_1's *being a Socrates-thought* is what is causally responsible.

What we just said about Bob applies to any entity, whether artificial or natural. Suppose M thinks *Socrates* and then *hit* and then *Plato*. Let m_1, m_2, and m_3 be the physical states realizing these thoughts. Unless m_1's being a Socrates-thought is what brings about m_2, and unless m_2's being a *hit*-thought is what brings about m_3, M hasn't thought *Socrates hit Plato.* It doesn't matter how reliable the mechanisms involved are. Of course, it is immaterial whether the engineer that designed M is a human being, or a Martian, or the invisible hand of evolution.

No matter how perfectly a physical system *simulates* thought, that system *doesn't* think *unless* what drives the relevant sequences of states are their *representational properties.* It doesn't matter whether the system is natural or artificial, whether it is the product of artifice or evolution.

Part IV. A compelling objection to our position is found in Jerry Fodor's work:

(CTM*)[109] Let $b_1 \ldots b_n$ be an arbitrary sequence of brain-states. You say that, if $b_1 \ldots b_n$ is to constitute a rational thought-process – indeed, if it is to constitute a thought-process at all – then a condition more or less like this must be fulfilled:

(#) for each i, b_i's *having such and such representational content* must be what brings about later entries in that series. It isn't enough that b_i do so.

But this simply isn't true. For the sake of argument, suppose that the condition described by (#) is *not* met. In other words, the sequence $b_1 \ldots b_n$ is *not* driven by those states' having such and such representational contents. Suppose that sequence is driven by the *non*-representational properties of the brain-states involved. Let us go further. Suppose that sequence is driven entirely by the *forms*, the *morphologies*, of those brain-states. In that case, we have a purely *form*-driven sequence of representations: a purely *formal* operation.

We know that *computations* are entirely form-driven operations. No consideration of meaning is involved. A computation is a kind of inference: a *formal*-inference. An inference is a thought-process. So for $b_1 \ldots b_n$ to

constitute a thought-process, it is *not* necessary that b_1's having such and such representational content have any causal role in brining about b_2. It is necessary only that b_1's formal -- its syntactico-morphological properties -- be involved in the right way.

So, in direct contradiction to what you say, the causal effectiveness of semantics (representational content) is *not* a necessary condition for the existence of thought-processes.[110] The very existence of computations – of purely formal thought processes -- shows as much.

A calculator is an environment in which representations follow one another in ways that are consistent with the demands of rationality. Further the behavior of a calculator is entirely form-driven: semantics plays no part. Syntax alone is operative – and syntax is all that is needed. So calculators *do* engage in formal thought and, therefore, in thought of some kind.

(CTM*) involves a confusion of two radically different meanings of the term "form". Once this confusion is exposed, the argument in question collapses.

The confusion is that between *syntactic* form and *physical* form.[111] These are two entirely different things. A form-driven operation is a computation only if the form in question is *syntactic* form. Physical form won't do. But a syntax-driven operation *is* a semantics-driven operation, contrary to what is generally thought. So, contrary to what CTM* says, if $b_1 \ldots b_n$ is to constitute a thought-process – whether "formal" or "informal" – it is *de rigueur* that b_i's having such and such representational content be

causally effective in bringing about later entries in that series.

In order to clarify the relevant issues, it will help if we momentarily leave aside brain-states, and instead discuss sentences. In due course, we will map what we say back onto brain-states.

First of all, *sentences* don't have physical form. The *sentence* "snow is white" is an abstract object. It can be instantiated by sounds, by inscriptions on a piece of paper, by hand-signals, or by smoke-signals. *Tokens* of that sentence have physical form. But the sentence *per se* does not.

Notice that tokens of that sentence can have very different physical forms. In terms of its physical form, a stream of sound has very little in common with a series of hand-movements or an inscription on a piece of paper. Even if we confine ourselves to written tokens of that sentence, any two tokens can have radically different forms. SNOW IS WHITE and *snow is white* have quite different physical forms.

But any two tokens of "snow is white" have the same *syntactic* form. A spoken token of that sentence has the same *syntactic* form as any written token of it. SNOW IS WHITE and *snow is white* both have the same syntactic form. They consist of a noun, followed by a one place-predicate.

It is clear what *physical* form is. The physical form of a sentence token is its shape – its "morphology". Obviously the *syntactic* form of a sentence-token isn't its morphology, since syntactically identical sentence tokens can have virtually no morphological properties in common (consider a pattern of hand-signals and a series of sounds). So what *is* syntactic form?

The syntactic form of a sentence is a function of the more generic facts about its semantics. Consider the following two sentences:

(a) "Jerry is a genius, and all geniuses smoke cigarettes."

(b) “Philmore is a moron, and all morons eat automobiles.”

These sentences are said to have the same syntactic form, namely:

(c) *so and so is a phi, and all phi’s stand in relation R to psi’s.*

(a) has (c) for its form because the constituents of (a) have certain semantic properties. “Jerry” is a singular term, “is a genius” is a predicate of some kind, “all” is a certain kind of quantifier (roughly, it denotes a certain quantity), “smoke” denotes a two-place relation, and that “cigarettes” is a noun of some kind. (a)’s syntactic structure is a reflection of its semantics. Exactly similar remarks apply to (b). (a) and (b) are syntactically the same because their semantic profiles are relevantly similar.

Nothing about semantics, or therefore about syntax, can be read off of the purely *morphological* facts about (a). A brilliant Martian who saw a token of (a) would have no reason to believe that (a) had (c) for its form. He would have no reason to believe such a thing *until* he knew that “Jerry” was a singular term, that “is a genius” is a predicate, and so forth. Morphology to no degree fixes semantics or, therefore, syntax. So, by itself, sensitivity to morphology does not qualify as sensitivity to semantics or, therefore, to syntax. If the term “form” is taken to denote *morphology*, then a purely “formal” operation is not a syntax-driven process. If the term “form” is taken to denote *syntax*, then a purely “formal” operation involves sensitivity to semantics. So contrary to what CTM* says, purely “formal” inferences *are* semantics-driven.

In this connection, it is worth bringing up a development in semantics. Until recently, the standard view is that syntax-defining terms e.g. “not”, “all”, and “for some”, do not denote anything. This is quite possibly false.

Consider the sentence

(1) “snow is not white”.

This is equivalent to

(2) “the sequence <snow is white> is such that its one member is false.”[112]

So the “not” in (1) is a second-level predicate; it picks out a property had by sequences of propositions (or propositional functions). So “not”[113] picks out the property of being such that its sole member is false. The same is true of “and”. Consider the sentence

(3) “snow is white and grass is green”.

This is equivalent to

(4) “the sequence <snow is white, grass is green> is such that both its members are true.

The “and” that occurs in (3) is converted, so to speak, into the predicate “is such that both its members are true.” So, in effect, “and” picks out the property of having two true members. Exactly similar remarks apply to “for some”, “for all”, “or”, and so on. So-called formal truth *does* turn on semantics; it turns on what is picked out by “and”, “or”, and the like. “Snow is white and grass is green” has a different syntax from “snow is white or grass is green”, but only because what “and” *picks out* is different from what

"or" picks out. Sensitivity to semantics is needed to register the syntactic, and logical, differences between those sentences. In general, syntactic differences turn out to be semantic; and syntactic differences are no more a matter of morphology than are any other semantic differences.

Consider the following sentence-token

(A) "Cornelius III is a pusillanimous ambulatory psychotic."

If that sentence is true, then so is

(B) "Cornelius III is pusillanimous."

But for every language in which the transition from (A) to (B) is logical, there are infinitely many in which it is *illogical*. There are infinitely many possible languages in which tokens morphologically identical with (A) mean *aardvarks never play checkers* and in which tokens morphologically identical with (B) mean *aardvarks always play checkers.*

By itself, morphology doesn't impose any constraints at all on logical form. So if we take into account only the morphologies of (A) and (B), and leave aside facts about their semantics, then (depending on one's views on the arithmetic of infinite quantities) the chances of the transition from (A) to (B) being logic-tracking are either vanishingly close to zero or they are no better than 50%. *Given* that (A) and (B) have certain semantic profiles, it follows that the transition in question is logic-tracking. So *given* the operative semantic background, that transition is logical. But from the viewpoint of a purely morphology-driven being, that semantic background is non-existent. So such a being would *not* be acting logically in transitioning from (A) to

(B).

Here is a widely held position.

Semantics can be “mirrored” in syntax. Thus, semantic relations like that of logical consequence are capable of a purely syntactic characterization. Syntactical facts are, in their turn, capable of a purely morphological characterization. Because of this, it is possible for purely morphology-driven operations to constitute logic-driven operations. Thus thought can be “mechanized” – it can be “morphologized”.[114]

There are two points to make here. First, in light of recent work in semantics, it is pretty clear that logical form *cannot* be adequately mirrored in syntactic form. So even if syntactical facts were capable of a purely morphological characterization, it would not follow that logical form was capable of such a characterization. Second, and more importantly, even if logical form *could* be perfectly mirrored in morphology, that would be irrelevant to the prospect of “morphologizing” (or otherwise “mechanizing”) thought.

Let us start with the second point. In English, sentences that seem to have the same syntax may diverge in logical form. “Bill is a green duck, therefore Bill is duck” is valid; “Bill is a decoy duck, therefore Bill is a duck” is not valid. For this reason, we are told that, in natural languages, syntactic identities conceal differences in logical form: in English, a given syntactic form can (supposedly) be associated with different logical forms, and there is thus no function from syntactic form to logical form.[115]

To simplify discussion, let us focus on languages in which this divergence

of logical and syntactical form doesn't occur. For the sake of argument, suppose that, in some language L, syntactic facts adequately reflect facts about logical form. So, in L, there a function from logical to syntactical form.

For the sake of argument, let us further suppose that, in L, there is a purely morphological characterization of syntactic form. There is a function from the morphologies of L sentences to their logical forms.

L is thus the kind of language that logicians dream about. In L, there is a purely syntactic characterization of logical form, and also a purely morphological characterization of syntactic form. In L, there is therefore a purely morphological characterization of logical form.

Here is where matters become problematic. Any morphological characterization of logical form will be *relative* to a language and, thus, to a semantic system. Let S1 and S2 be two sentence-tokens of L such that S1 entails S2. To fix our ideas, let us suppose that, in L, S1 means *grass is green* and S2 means *something is green.* In L, a sequence consisting of S1 followed by S2 is logic-tracking, and, for the reasons described a moment ago, that fact will be reflected in the morphology of that sequence. (Let S1-S2 be our shorthand for that sequence.)

But there are infinitely many languages such that, relative to any one of them, a sequence morphologically identical with S1-S2 is logically false or altogether meaningless. Let L* be any such language. The difference between a case where that transition is logical and one where it isn't has to do entirely with the background semantics. If L is operative, the transition is logic-tracking. If L* is operative, that transition is not logic-tracking.

By itself, the morphology involved is as consistent with the operation's being fallacious (or nonsensical) as it is with that operation's being logic-tracking. The difference between a sequence that tracks logic and one that doesn't lies in the semantics involved. If the right semantic background is

absent, it doesn't matter if S1 leads to S2. What must be operative, therefore, is S1's having a certain semantics or representational content. The sequence S1-S2 is logical relative to L, but illogical relative to infinitely many other languages. Given only the morphology of that sequence, it is just as much an expression of L* as it is of L. If that sequence is to constitute an expression of L, that means It is not enough that *S1* cause S2. If the sequence S1-S2 is to be logic-tracking, that means that a certain semantic system – a certain assignment of meanings to S1 and S2 – must be operative. So if S1-S2 is to be logic-tracking, it is necessary that S1's *having a certain semantics* be operative.

closing the argument

Let us sum up. Let B_1 be a physical-state realizing a belief that P and let B_2 be a physical-state realizing a thought that Q. If the sequence B_1-B_2 is to constitute a thought-process, it is not enough that B_2 cause B_2 via a reliable mechanism. It is necessary also that B_1's *having representational content* be causally involved. When content is inert, there is no thought.

(CTM*) responds to this by saying:

"False! B_1's having that representational content needn't be causally involved. The sequence B_1-B_2 could be a purely form-driven – a purely morphology-driven – operation; and in that case, it would qualify as a formal-inference, and thus as a cognitive achievement. After all, a formal-inference is one that is *not* driven by anything's having such and such representational content. So even if B_1's having a certain representational content is completely inert, the sequence B_1-B_2 can qualify as a formal inference (a computation)."

But a so-called formal inference *is* semantics-driven. If the sequence B_1-B_2 is to constitute a formal inference, B_1's having such and such representational content absolutely must be involved. Otherwise we don't have anything logic-evaluable. Indeed, we don't have a thought-process at all.

References

Barwise, Jon and Perry, John. 1999. *Situations and Attitudes.* Palo Alto: CSLI Press.

Davidson, Donald. 1970. "Mental Events." In *Experience and Theory.* (Lawrence Foster and John Swanson eds.). University of Massachusetts Press and Duckworth. Reprinted in *Essays on Actions and Events*, pp. 207-225. 1984. New York: Oxford University Press.
-- 1973. "Freedom to Act." In *Essays on Freedom of Action.* (Ted Honderich, ed.). Routledge and Kegan Paul. Reprinted in *Essays on Actions and Events*, pp. 63-81. 1984. New York: Oxford University Press.

Fodor, Jerry. 1968. *Psychological Explanation.* New York: Random House.
-- "Searle on What Only Brains Can Do" and "Yin and Yang in the Chinese Room". *Behavioral and Brain Sciences 3 (3):* 431-432 and 451-452. Reprinted in *The Nature of Mind* (Davis M. Rosenthal, ed.) pp. 520-521, 524-

525. New York: Oxford University Press.
--1981a. *Representations.* Cambridge: The MIT Press.
--1981b. “Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology”. In *Representations*, pp. 225-253. Cambridge: The MIT Press.
--1987. *Psychosemantics.* Cambridge: The MIT Press.
-- 1990a “Fodor’s Guide to Mental Representation”. In *A Theory of Content and Other Essays.* Cambridge. The M.I.T. Press.
-- 1990b. “A Theory of Content I”. In *A Theory of Content and Other Essays.* Cambridge. The M.I.T. Press.
--1990c. “A Theory of Content II”. In *A Theory of Content and Other Essays.* Cambridge. The M.I.T. Press.
--1994. “Fodor”. In *A Companion to the Philosophy of Mind.*(Samuel Guttenplan ed.). Cambridge: Blackwell.

Edmund Gettier: 1963. “*Is Justified True Belief Knowledge?*” In *Analysis*, v. 23.

Jackson, Frank and Pettit, Philip. 2004. “Causation in the Philosophy of Mind.” In *Mind, Morality, and Explanation*, pp. 45-68. New York: Oxford University Press.

Kim, Jaegwon. 1993. *Supervenience and Mind.* Cambridge: Cambridge University Press.

Lewis, David. 1972. “Psychophysical and Theoretical Identifications.” *Australasian Journal of Philosophy* L, 3 (December 1972): 249-258. Reprinted in *The Nature of Mind* (Davis M. Rosenthal, ed.), pp. 204-210.

New York: Oxford University Press.

McCawley, James D. 1981. *Everything that Linguists have always wanted to know about Logic.* Chicago: University of Chicago Press.

McGinn, Colin. 1989. *Mental Content.* Oxford: Basil Blackwell.

Putnam, Hilary. 1967. “The Nature of Mental States”. Originally published as “Psychological Predicates”. In *Art, Mind, and Religion.* (W.H. Capitan and D. D. Merill, eds.), pp. 37-48. Reprinted in Reprinted in *The Nature of Mind* (Davis M. Rosenthal, ed.), pp. 197-203. New York: Oxford University Press.
--

Russell, Bertrand. 1905. “On Denoting.” Reprinted in *The Philosophy of Language.* A.P. Martinich (ed.), pp. 203-211. Oxford: Oxford University Press.

Searle, John. R. (1980) “Minds, Brains, and Programs.” *Behavioral and Brain Sciences 3 (3): 417-457.* Reprinted in *The Nature of Mind* (Davis M. Rosenthal, ed.), pp. 509-519. New York: Oxford University Press.

Davidson's Misconceptions about Rationality

Abstract: Alan Turing advocated a kind of functionalism: A machine M is a thinker provided that it responds in certain ways to certain inputs. Davidson argues that Turing's functionalism is inconsistent with a certain kind of epistemic externalism, and is therefore false. In Davidson's view, concepts *consist* of causal liasons of a certain kind between subject and object. Turing's machine doesn't have the right kinds of causal liasons to its environment. Therefore it doesn't have concepts. Therefore it doesn't think. I argue that this reasoning is entirely fallacious. It is true that, in *some* cases, a causal liason between oneself and an object is *part* of one's concept of that object. Consequently, to grasp *certain* propositions, one must have certain kinds of causal ties to one's environment. But this means, at most, that we must rethink some old views on what rationality is. It does *not* mean, *pace* Davidson, that a precondition for being rational is being causally embedded in one's environment in a certain way. If Turing's machine isn't capable of thinking (I leave it open whether it is or is not), that has nothing to do with its lacking certain kinds of causal connections to the environment. The larger significance of our discussion is this: rationality consists *either* in one's ability to see the bearing of purely *existential* propositions on one another *or* rationality is simply not to be understood as the ability see the bearing that propositions have on one another.

1. In this paper, I'd like to make a narrow point and a broad point. The narrow point is that what Donald Davidson says about Turing machines – specifically, his argument for holding that they don't think – doesn't go through.[116] The broad point is: Given that these reasons of Davidson's

don't go through, there is some reason to believe that *one* of the following two claims holds.

(a) Intelligence is *not* to be understood as the ability to see the bearing that propositions have on other propositions.

(b) Intelligence is to be defined as the ability to see the bearing of *purely existential* propositions on other purely existential propositions.

Of course, (a) and (b) cannot both hold.

2. First some background. Putnam showed that one's concepts of objects – one's abilities to have thoughts about objects -- often have a causal *component.* At least a *part* of one's concept of water or of Bob consists in there being some kind of causal connection between oneself and those things. [117]

Putnam showed this through thought-experiments like this one. Twin-John and John are molecule for molecule duplicates of each other. So John and Twin-John are exactly alike *leaving aside* facts about the (distal) causes of their conditions. Given this, suppose that John has a visual sensation that is caused by Bob, and that twin-John has a qualitatively identical visual sensation that is caused by twin-Bob. John is perceiving *Bob*, not twin-Bob. And twin-John is perceiving twin-Bob, not Bob. From his perception, John thus 'uploads' a concept of Bob, not of twin-Bob. And from *his* perception, twin-John uploads a concept of twin-Bob, and not of Bob.[118]

The fact that John has a concept of Bob, and not twin-Bob, derives from the fact that John has a causal connection to Bob that he doesn't have to twin-Bob. And the same thing *mutatis mutandis* explains why twin-John has a

concept of twin-Bob, and not of Bob.

Thus John has thoughts about Bob *in virtue of* (at least in part) his having a certain causal connection to Bob. Thus John's concept of Bob – his ability to have thoughts about Bob – consists, at least in part, in his having a certain causal connection to Bob. (The same thing is true *mutatis mutandis* of twin-John and twin-Bob.) In this way, Putnam (successfully) shows that a causal connection to an object may be, at the very least, a *part* of one's concept of that object.

In many publications, Jerry Fodor says our concepts of objects *consist* in causal relations of a certain kind between ourselves and objects.[119] As we'll see in a moment, Donald Davidson takes the same view.[120]

I now want to show that this view of Davidson's seriously lead him astray in his analysis of a famous problem posed by Alan Turing.

3. Turing posed the following problem. Let M be some machine, and suppose that M responds to inputs in exactly the way that a thinking being would respond to those inputs.[121] So if you say to M 'should we go back to the gold-standard?' or 'what did you think of the game last night?', M gives you just the kind of answer that a cognitively normal human gives you. And, of course, the same is true for any other question or assertion that you might direct to M. Turing asks: under that circumstance, does M qualify as a thinker? Turing's implicit answer is 'yes'.

Turing is not, I think, making an epistemological point. He is not saying that, since there is as much data to support the view that the machine is thinking as there is to support the view that the person is thinking, there is *good reason* to hold that the machine thinks; he is not saying the machine's behavior is merely *evidence* of thought. (For it is obvious and trivial to say

that if M acts just like a person that is *evidence* – though, perhaps, defeasible evidence -- that he has the mental capacities of a person.). Turing is saying, it seems, that if the machine behaves indistinguishably from the thinking person, then *ipso facto* the machine is a thinker.

Davidson thinks that Turing is wrong. Now I myself *agree* with Davidson that Turing is wrong; I don't think that M thinks. (More specifically, I don't think that M qualifies as a thinker *merely* in virtue of behaving in the right way in response to certain inputs. M might qualify as a thinker for some other reason.) But my purpose here is not to discuss whether Turing is right or wrong. It is to discuss Davidson's criticisms of Turing. I submit that Davidson has given us no reason to think Turing wrong.

4. Davidson's criticism is this. Let John be some normal human being whose thoughts, and therefore whose sentence-tokenings, are causally 'hooked up' to external objects in the right kind of way. So John has de re thoughts about Bush, Socrates, light and water; and he expresses these when uses sentences containing the terms 'Bush', 'Socrates', etc.

Now let us suppose that we assemble some being that behaves just like John in terms its verbal and behavioral responses to inputs – externally, it responds in the same way as John to the sound 'was Plato more conservative than Aristotle?' and 'is there an even prime?' Let Robo-John be this artificial John-proxy. Of course John actually *means* Plato by 'Plato', actually *means* water by 'water', and so on. This is not (merely) because John is possessed of some mysterious subjectivity or mentality; it is (at least in part) because John is *causally* connected in the right kind of way to Plato and water, and his verbal behavior derives from this connectedness. John's current psychological condition is the fall-out, at least in part, of his being embedded in a certain physical environment in a certain way: and it is this

embeddedness which enables him to think about Plato and thus to *mean* Plato by 'Plato'. This, in effect, is what we learn from the Putnamian thought experiment described above.

Now Robo-John was put together in some laboratory; so *his* psychological condition is an artifact; it is definitely *not* the fall-out, even in part, of his being embedded in a certain environment in a certain way. (Of course, one might say: 'well, the scientists who put Robo-John together knew about Plato and water, and so on, and *their* ability to thinks about things was incorporated into Robo-John.' But we can easily side-step this complication by supposing that Robo-John wasn't created in a laboratory, but rather came together as a result of a sandstorm or some random quantum event.[122]) In any case, his relation to his environment is definitely *not* such as to underwrite his having concepts of Plato or water. So Robo-John *just cannot think about Plato or water or Bush.* He acts just like John; and he says (or rather mouths) the same things. But when he says 'Plato was wise', he isn't *referring to Plato.* He is shooting a blank; he cannot *mean* Plato by 'Plato', for he hasn't got the right causal connection to the man. And, Davidson says, this point applies mutatis mutandis to anything else Robo-John says or does.

Davidson concludes that Robo-John *does not think*: for he hasn't got the causal connections to his environment to underwrite the having of the concepts needed to think. Thus, even though Robo-John is indistinguishable (externally) from John, Robo-John doesn't think, while John does think. And Turing is therefore wrong.

5. Again I agree with Davidson that Turing is wrong. But Davidson's argument has little force against Turing. A brain in a vat cannot think about Socrates or about water. But it can still think.[123] When you dream about

unicorns and goblins, your dream-images are images of *nothing*; and the narrative of your dream represents nothing real.[124] But in dreaming – in experiencing this play of dream images – you are still thinking. Whatever the line is between thinking and non-thinking beings – between being a rock and being a being a Shakespeare -- you are, in dreaming, well on the right side of it. Obviously a brain in a vat can dream. What it cannot do is have thoughts that are object-involving with respect to Socrates or water. But clearly *something* mental – indeed, something *cognitive* -- can still go on in it.

Turing's machine may or may not be thinker: I leave it open whether it is. But if it *isn't* a thinker, that has nothing to do with its lacking the right kinds of causal connections. The fact that there are *certain* things Turing's machine cannot think about – e.g. Socrates – has to do with its not having the right causal connections. But if Turing's machine cannot think *at all*, that is not to be explained by saying that it doesn't have the right causal connections. What led Davidson astray here is his holding that concepts *consist* in causal connections with the external world, rather than merely *involving* them.

6. Of course, the following objection will be made:

To think is to entertain propositions or, at any rate, is to handle information. Turing's machine – and its cousins: the subject of a Cartesian nightmare, the brain in a vat – just cannot grasp things; it can't grasp the constituents of propositions (water, Plato, redness….). Therefore it can't grasp propositions, and thus cannot think at all.

There are a number of things to say in response to this. First of all, when you dream of griffins or unicorns, it certainly seems that you are grasping a great many *existential* propositions, even though you are not grasping any

object-involving ones. Suppose you are dreaming of a non-existent pink elephant. Now given somebody who was in a phenomenologically identical state who *was* causally connected to a pink elephant in a certain way, that person *would* be grasping an object-dependent proposition[125] that you are not. In the cognitive 'place', to so speak, where that person is grasping an object-dependent proposition, you are grasping nothing.[126] But you are still grasping various existential propositions: *there is a four-legged creature moving about in such and a such a way…*

Now one might *counter*-respond by saying:

> The constituents of these existential propositions are such that your grasping them consists, at least in part, in your standing in certain causal relations to the external world. Consider, for example, the existential proposition: there is a four-legged creature moving about…Surely your concept of a leg is an object-involving one, the same being true of concept of a creature and perhaps even of your concept of movement. Indeed, So once you take away all of a causal liaisons, you take away its ability to entertain even existential propositions like *there is a four-legged creature*.

I find this highly dubious. A brain in a vat, it seems to me, has a concept of a leg, of a creature, of movement, as much as anyone else, notwithstanding that the kinds of connections it has to the outer world don't sustain de re awareness of any of its constituents.[127]

Suppose you discovered – let us set aside, as irrelevant, *how* you discover this – that you are a brain in a vat. Your Plato-concept will thus turn out to be empty, and so will your Nathan Salmon-concept. But your leg-concept? Your movement-concept? Would you then think to yourself: "Gosh, I guess my

leg-concept$_O$ turned out to be empty after all, just like my Socrates-concept"? No – *leg* is a functionally defined concept. You don't have to be causally connected to an actual leg to grasp the concept of a leg; you just have to know what functional role legs have. And it is very hard to see how your brain-twin could lack such knowledge.

7. What is true of one's concept of Plato or of water seems not to apply to *all* of one's other concepts, but only some of them. Suppose there is a brain in a vat that is just like your brain *modulo* those properties of you (and your brain) that supervene on your *not* being a brain in a vat. Further, suppose that, one day, you put that brain in a body – so it can speak, move, and so forth. That creature will produce sounds like 'that creature is moving rather quickly'. Is that creature really 'shooting blanks' when it says such things? To be sure, where attempts to make *certain* statements are concerned, it will be shooting blanks. If it says 'Socrates drank hemlock' it will not mean by that sentence what you and I mean by it. But a disembodied brain in a vat can have an appreciation of what it is for an object to move from one place to another, of what sentience is, of what a leg does for a sentient creature.

Also that creature would surely have just as much a grip on mathematical and purely rational truths as you. It would be nonsense to say that, in virtue of the fact that you had certain causal liaisons to the external environment, you were a better mathematician than your recently embodied brain-twin.

8. There is no denying that the behavior of the recently embodied brain-twin is *replete* with mentality. And it is hard to deny that some of this mentality is cognitive and involves intelligence. Surely your brain-twin is not *stupid*; he is obviously not in the same category as something inanimate. Surely it is as intelligent as you, notwithstanding that it doesn't have a grasp

of Plato or Socrates.

But suppose that, despite all this, someone taking Davidson's line digs in his heels and insists that one cannot grasp any propositions without having certain kinds of causal liaisons to one's environment – without being (or having been) embedded in a certain physical environment in a certain way. (The idea would be, perhaps, that such causal liaisons so totally infect our cognitive lives that *nothing* proposition-bearing would be left over if one were stripped of these liaisons.)

I would suggest that, if that is really the case – and maybe it is -- then we must define intelligence *not* in terms of seeing relations among *propositions*, but in some other way. For it seems to me quite obvious that a brain in a vat and the victim of a Cartesian nightmare can have all the *intelligence* in the world; Einstein's disembodied brain-twin is surely not *stupid*, even though he will shoot a lot of conceptual blanks. In other word, where the real Einstein -- who *is* appropriately embedded in a physical and social environment --*will* be entertaining various object-dependent thoughts, Einstein's brain-twin will (at least according to Evans) be entertaining thought-*like* things that have gaps in them, due to its lacking the right causal connections to the outside world. But in virtue of having these pseudo-thoughts, or "empty" thoughts, Einstein's brain-twin still obviously has *intelligence.* Einstein's brain-twin is not stupid: it is obviously not in the same category as a rock or a cactus.

So *if* one needs to be embedded in a physical environment in a certain way to grasp any propositions at all, *then* what that means is that intelligence, and cognitive capacity, are not to be defined in terms of proposition-manipulating ability. It *doesn't* mean that a thing, like Turing's machine, which is not appropriately embedded fails to think.

9. It is, of course, strange to say that intelligence consists in something

other than the ability to the bearing that propositions have on other propositions But such a view is not without a foundation; it is, I think, the distillation of epistemic externalism.

Kripke (1977) made a very compelling case that one can assent to a proposition *and* to its negation without being irrational. Pierre believes the proposition expressed by 'London is not lovely' and also that expressed by 'London est jolie'. And this is not because Pierre lacks acumen. It isn't that he has failed to deduce some proposition from some other proposition that he accepts. Pierre is not irrational; he isn't even guilty of logical laziness – of not making the right inferences from what he already knows. (If we *stipulated* that Pierre was logically omniscient, he would *still* be in the predicament that Kripke describes.) One lesson that might be drawn from this is: rationality (intelligence) is not to be defined in terms of what one does with the propositions one grasps. This has a certain similarity to the idea that intelligence (rationality) is not be defined in terms of proposition-manipulating ability.

Actually it seems that this last point – intelligence is *not* to be defined as the ability to manipulate *propositions* – is, in some respect, the essence of externalism.

In any case, it seems plain that a disembodied brain in a vat can have plenty of *intelligence*, even though it has no de re awareness of anything external, and is therefore incapable of grasping many propositions that you and I can grasp. So *if* Turing's machine lacks intelligence, that is not because it doesn't have right kind of causal connection to the outside world; it would be for some totally different reason. Its lack of causal connections deprives it of certain *concepts* but not of intelligence. So Davidson' has by no means proven that Turing's machine cannot think; for given only what Davidson says, it is still an open question whether Turing's machine is intelligent. Thus

Davidson's attack on Turing is misguided.

10. I should qualify something said earlier. We've seen that Davidson's attack on Turing doesn't go through. We've also seen that this fact gives *some* credence to a conception of intelligence (or rationality) that is different from the traditional one. (The traditional one is: intelligence consists in some kind of *proposition*-manipulating ability.) But the spuriousness of Davidson's attack by no means *proves* that intelligence is not to be thought of the traditional way. As I suggested earlier, whenever one is grasping some *object-dependent* proposition, one is also, it seems, grasping many existential propositions. When I see Toonces the cat chasing Fido the dog, I am aware of some object-dependent fact, namely *Toonces is chasing Fido*. But necessarily involved in this grasp is a grasp of various existential propositions. I don't *just* see Toonces; perception, and representational mental activity generally, is predicational; I must see Toonces *as* having certain properties (as having fur of a certain color, as being four-legged, as being a certain distance from me, and so on). So my visual awareness of Toonces seems to mediated by propositions (or, in any case, by information) that is existential (*there is a four-legged creature with thus and such properties…running after a four-legged creature with such and such properties*). Now notice that, where the proposition *Toonces is chasing Fido* is *object-involving* with respect to a certain cat and a certain dog, the just mentioned existential propositions are not thus object-involving: *there is a creature with pointy ears…chasing a creature with floppy ears.*…is not object-involving, at least not with respect to either Toonces or Fido. Basically, 'existentializing' a proposition tends to get *rid* of the objects in that proposition; by replacing *Toonces* with a variable, and then prefixing the resulting open proposition with an existential quantifier, one turns a proposition that is object-involving with respect to a

certain thing into one that is not object-involving with respect to that thing. It thus seems that if an object-involving proposition could be *completely* 'existentialized' (i.e. if each spatio-temporal object implicated in that proposition could be subject to the procedure we just described), the result would, of course, be a completely object-independent proposition. And a *completely* object-independent proposition is one the grasping of which *does* supervenes on one's brain-states considered apart form their causal origins (or, if you prefer, if on one's mental states narrowly individuated). To grasp the proposition

(*) *Socrates was bald*,

it is not enough that my brain be in a certain state; my brain-state must have certain causal origins. But that is entirely because (*) is *object*-involving; it has Socrates himself as a constituent. Obviously given a proposition that didn't have Socrates as a constituent, one could grasp that proposition without having a causal connection to Socrates specifically. By analogous reasoning, given a proposition that doesn't spatio-temporal object O as a constituent, one can grasp that proposition without having a causal connection to O specifically. So if a proposition is *completely* object-independent, then one can grasp it, in principle, without having a causal connection to *anything*: one's grasping it would supervene on one's brain-states (or mental states) considered *apart* from their causal pedigrees. So *purely* existential propositions can be grasped by a brain in a vat. (Whether *there are* such things as purely existential propositions is another question. I think there are. Presumably, the principles of mathematics and logic don't have spatiotemporal individuals or kinds for their constituents – though Putnam, I believe, has challenged this orthodoxy.) Given someone who

believes that Turing's machine is intelligent (i.e. thinks and does so competently) and who *also* believes that intelligence is to be analyzed as proposition-manipulating ability, that person's best bet , perhaps, would be to say that Turing's machine operates on purely existential propositions (propositions that do not have spatio-temporal objects or kinds for constituents).

References

Davidson, Donald. 2004. 'Turing's Test' in *Problems of Rationality.* Oxford: Clarendon Press.

Kripke, Saul. 1979. "A Puzzle about Belief". In *Propositional Attitudes.* Nathan Salmon and Scott Soames (eds.). Oxford: Oxford University Press.

McGinn, Colin. 1988. *Mental Content.* Oxford: Basil Blackwell.

Putnam, Hilary. 1975. 'The Meaning of "Meaning"' in *Collected Papers vol. III.* Cambridge: Harvard University Press.

Jerry Fodor's Pseudo-computationalism

There is good reason to believe that mental entities are identical with, or realized by, neuro-cerebral entities. If this is right, then your being in pain or thinking about the number two is realized by some brain-event or series of such events.[128] [129]

Brains are obviously governed by the laws of physics – just like billiard balls and gases. At the same time, our thought processes are obviously *rational*, at least up to a point. A person who believes *Bill is either in Greenland or Iceland, and he is not in Iceland* is likely to form the belief that Bill is in Greenland. To be sure, we are not *always* rational. But thoughts and, more generally, mental contents succeed one another in ways that are, to a high degree, consistent with the demands of rationality.[130]

There is a problem. Brains are governed entirely by the laws of physics. The laws of physics are purely descriptive, not prescriptive. In moving this way as opposed to that, the billiard ball isn't doing anything right or wrong, it isn't being rational or irrational. Physics alone decides what the billiard ball does. Consciousness of norms – or of canons of good reasoning – plays no part. Our brains are governed by the laws of physics no less than billiard balls. The laws of physics alone decide how they will behave. There is no room for awareness of the canons of good reasoning to play a part. And our mental lives are entirely realized by our brain activity.[131] But our mental activity obviously *conforms* to such canons, at least to a high degree.[132]

So we are left with a problem. Our mental activity is rational, i.e. it

proceeds in ways that are countenanced by canons of good reasoning. At the same time, our brains, and therefore our minds, are governed entirely by the laws of physics. The laws of physics are not laws of rationality. So laws of rationality play no part in our thinking.

There is an elegant solution to this problem. This solution constitutes one of the more compelling arguments for the *computational theory of mind* (ctm). This solution is most clearly and forcefully articulated by Jerry Fodor.

The purpose of this paper is to show that this solution involves a massive confusion, and is therefore no solution at all. Thus at least *one* of the arguments for CTM, or at least for one version of it, proves to be quite fallacious. I leave it open whether it nullifies CTM or merely removes one of its supporting pillars.[133]

Fodor's solution

From one viewpoint, a symbol is a purely physical entity: a noise, a deposit of ink, a pattern of light on a monitor. Being physical entities, symbols (or symbol-tokens, strictly speaking) obey the laws of physics; their interactions are entirely physics-driven. To explain those interactions, it isn't necessary to invoke anything other than physical entities and physical laws.

But it is easy to create an environment in which the interactions among symbols "track" the laws of logic. Such environments are created and exploited everyday. They are called "computers".[134]

Consider the symbols "P", "Q", and "if P then Q". Those symbols have various physical properties. Because they have those properties, the laws of physics dictate that they act in certain ways in certain environments. We can construct an environment in which those symbols interact in the *right* way: an

environment in which “P” and “if P then Q” is followed by “Q”. We can construct environments in which symbols follow on another in a way consistent with the demands of rationality. Such environments are entirely physical; and the behavior of the symbols in question is entirely physics-driven. But the structure of the environment guarantees that those entirely physics-driven processes conform to the demands of logic. If the input is “(P or Q) and not Q”, the output will be “P”. But the process is entirely physics driven: through clever engineering we have ensured that, given the physics involved, output will be logically appropriate to input.

We have thus dissolved the paradox earlier described. There doesn’t seem to be any difficulty creating an environment in which entirely physics-driven processes respect the canons of logic – in which the descriptive and the prescriptive coalesce.

But there is an obvious response:

It is easy enough to create a purely physical environment which *mimics* or *simulates* thought: an environment in which “P” and “if P then Q” are followed “Q”. But that is not enough for the existence of thought or inference. To infer “Q” from “P” and “if P then Q” is to recognize that the latter two symbols stand in a certain logical relation to the first. The essence of inference is *recognition* of some bearing relation: a relation of entailment or probabilification. Let us tell a story. You believe P; you can also believe (if P then Q); and these beliefs *cause* you to believe Q. Have you made an inference? Not necessarily. It is not enough that the first two beliefs bring about the third: they must bring it about in the right way: they must do so by way of your recognizing the existence of some bearing relation. Now let us tell a different story. You believe P and also (if P then Q). You also recognize that Q is entailed by P, conjoined with (if P then Q). These beliefs conspire

with this recognition to make you believe Q. *Here* we have an inference. To have a rational thought-process, it is not enough for symbols to follow one another in a certain order. It is also necessary that their falling into that order embody an awareness of bearing-relations holding among them. In a computer, the symbols follow one another in the right order; but their doing so doesn't involve a recognition of any bearing relations holding among those symbols. So the computer doesn't think at all.

Fodor has a clever response to this kind of objection. What is a computation? It is an inference: but not just *any* kind of inference. If you infer "Fred is an animal" from "Fred is an elephant", you have made an inference; indeed, you have made a correct inference. But you have not performed a computation. Why not? Because the inference in question was not purely *formal*; it involved consideration of the *semantics*, and not just the *syntax*, of the symbols involved. A computation is a purely *formal* inference; it is an entirely syntax-driven inference.[135] Consideration of semantics can play no part.

From a purely formal or syntactical viewpoint,

(i) "Fred is an elephant, therefore Fred is an animal"

is indistinguishable from

(ii) "Fred is an elephant, therefore Fred is a smoker".

Of course, (i) is a valid inference, whereas (ii) is not. But to see this one must know the *meanings* of the symbols involved. So (i), though valid, is not a formal inference. A purely syntax-driven transition from “Fred is an elephant” to “Fred is an animal” is not valid.
Thus, it is no computation.

Matters are different with:

(iii) “Fred is an elephant; therefore Fred is either an elephant or an aardvark”

To know that (iii) is true, you don’t have to know what is meant by “elephant” or “aardvark” or “Fred”. (iii) is formally true. A purely syntax-driven transition from “Fred is an elephant” to “Fred is either an elephant or an aardvark” constitutes a valid inference. It is a computation.

The syntax of an expression is its *form*. Ultimately, form reduces to morphology.[136] Morphology is shape. Shape is entirely physical. A morphology-driven process is an entirely physics-driven process. But some morphology-driven processes are also *inferences*: they are *formal* inferences; they are computations. So entirely physics-driven processes can qualify as inferences: they can qualify as things which can meaningfully and even correctly be described in normative terms – as right or wrong, justified or unjustified.

So it is easy to dissolve the paradox mentioned above. We advocate CTM: we assume that brains (or brains-cum-bodies) are environments that house morphology-driven interactions among symbols. Such interactions are purely physics-driven, but they also qualify as inferences of a certain kind: formal inferences – inferences that embody sensitivity to syntactic, and thus to

logical, form. Because of the tight relation between an expression's logical form and its syntax, and between its syntax and its morphology, it follows that entirely morphology-driven operations can constitute inferences, and thus cognitive achievements, of a certain kind. So CTM dissolves the paradox mentioned a while ago.[137] [138]

the purpose of this paper

I wish to show that the line of thought just presented is entirely spurious. There are many problems with it. One problem is that it involves a confusion of two entirely different meanings of the word "form". There is form in the syntactic sense, and there is form in the *physical* sense. These two things have practically nothing to do with each other. Any connections between them are exceedingly indirect. Two physical entities that have exactly the same syntactic form can have nothing in common at the level of morphology. And two morphologically identical physical entities can have nothing in common at the level of syntax. So morphology-driven operations, though "formal" in one sense of the word, are by no means syntax-driven. Genuinely syntax-driven operations can *involve* sensitivity to morphology (in the sense of physical shape[139]); but, as we will see, sensitivity to morphology plays a decidedly subordinate role in genuinely syntax-driven operations. And it is, at best, extremely misleading to say that syntax-driven operations can be identified, even to a first approximation, with morphology-driven ones.

Further, a syntax-driven process *is* a semantics-driven process. Any inference from "Fred is an elephant" to "Fred is an elephant or Fred is an aardvark" is replete with knowledge of semantics – hardly less so than the

inference from “Fred is an elephant” to “Fred is an animal”. A syntax-driven operation can no more be identified with a morphology-driven operation than can a semantics-driven operation.

Syntax, we will see, is an aspect of semantics. The syntax of a sentence (or sentence-token) is identical with the more generic facts about its semantics.[140] Semantics consists in word-world relations. For a being to engage in syntax-driven operations, it must have semantic knowledge; and this involves having extra-linguistic knowledge. The ability to engage in syntax-driven operations thus presupposes a wealth of knowledge and, therefore, of cognitive and representational abilities. So the ability to engage in syntax-driven operations cannot possibly have a cognitively foundational role.[141]

There is another, altogether different problem to consider. Syntactic form does *not* typically track logical form, contrary to what used to be thought. Recent work in logic and semantics has demonstrated this.[142] Nearly all sentences have a context-sensitive component; this fact creates a chasm between syntactic and logical form, and it makes it impossible for both semantics and also logical form to be adequately mirrored by syntax. This, in turn, makes semantic, and extra-semantic, knowledge a pre-requisite for being able to engage in operations that track, or embody any sensitivity to, logical form.

For these reasons, the idea that computing could play a foundational role in thought is simply confused. In fact, computing is an extremely special and derivative form of thought. Granted, there is *a* sense in which a being lacking semantic, or extra-semantic, knowledge can “compute”. But that sense of the word “compute” is entirely irrelevant to attempts to explain the foundations of thought. Failure to distinguish real computing, which presupposes a wealth

of cognitive abilities, from faux-computing, which does not, is abetted by a tendency to wrongly identify form in the syntactic sense with form in the morphological sense.

syntax versus semantics

Consider the sentence:

(iv) “If Fred is an elephant, then Fred is either an elephant or Fred is a chipmunk”.

Supposedly, this sentence is *formally* true. But what does that mean? Does it mean that its truth has nothing to do with its semantics? That would be nonsense.

Let us start with the obvious. The symbol “or” (or “either…or…”) could mean anything; it could mean *and* (or *both…and…*); it could mean *because*; it could refer to Plato; or it could mean nothing at all.

To know that (iv) is true, one must know the semantics of “or”; one must know what it means. One must know that it doesn’t refer to Plato or denote the relation of cause and effect. So to recognize that (iv) is “formally” true, one must have knowledge of semantics.

Some expressions are said to be syntax-defining, while others are not. Sometimes the former (or, at least, a certain subset of the former) are referred to as “logical particles”. These include “and”, “if…then…”, “not”. Given a sentence containing one of these terms, if you replace that term with some

other term, you have changed the syntactic structure of that sentence. Consider (iii). Suppose you replace the “or” with an “and” or a “because”. In that case, you have changed the syntactic structure of the sentence. Further, you will have turned a sentence that is logically true into one that is not logically true – into one that is, in fact, logically false.[143]

Expressions like “Fred” and “elephant” are said *not* to be syntax-defining. Suppose that, in (iv), you (uniformly) replace “Fred” with “Jerry”. In so doing, you have *not* changed the syntax of the sentence: and the sentence you produce will be logically true exactly if the original sentence had that property. Apparently, the property of logical truth is preserved under (uniform) substitutions of non-logical expressions for non-logical expressions.

Non-logical expressions like “Fred”, “Jerry”, and “elephant” denote individuals or properties. Logical particles (e.g. “or”, “for some”) are typically thought *not* to denote anything; they are said to have a purely “formal” or “syntactic” function. There are two points to make in response to this. First, even if they do not denote anything, one still has to know their semantics to know that sentences like (iv) are logically (or formally) true. You could be the most brilliant logician in the world. If you don’t know what is meant by “or” or “therefore”, you simply won’t have any way of knowing that (iv) is logically true. And there is no way to read the semantics of such expressions off of their morphology. An expression with the same semantics as “or” could have any morphology at all. Let L be a language just like English except that the L-word for “or” is “and” and the L-word for “and” is “or”. In L, (iv) is logically *false*, not logically true. Let L* be a language just like English except that, in L*, “or” means “banana”. In L*, (iv) is meaningless; it isn’t true or false.

So even if syntax-defining terms are non-referential, one must still know

a great deal about semantics to determine if a sentence is logically true or not.

There is another point to make. A case can be made that syntax-defining terms *do* refer. I myself have no doubt that this is so; and it is the view of many logicians and linguists. Such terms can be seen as referring to second-level functions or properties. Consider the sentence:

(v) Sally is a square or Sally is not a square.

This is logically, and perhaps semantically, identical with:

(vi) Or*<Sally is a square, Sally is not a square>,

where "or*" denotes a function that assigns truth a sequence of sentences exactly if at least one of them is true. So "or*" can be seen as denoting the property of being a sequence of sentences such that at least member of that sequence is true. (vi) is obviously logically equivalent with (v). So the "or" in (v) can be seen as a variant of "or*"; it can be seen as denoting the second-level function (or property) just described.[144]

Exactly similar remarks apply to all syntax-defining terms. Consider the sentence:

(vii) "snow is not green".

This is equivalent to

(viii) Not*<snow is green>,

where “not*” denotes a function that assigns truth to a sequence exactly if that sequence comprises one sentence and that sentence is false. By reasoning similar to that given a moment ago, the “not” in (vii) can be seen as denoting that second-level function. On similar grounds, Frege argued that quantifiers denote second level functions.[145] (“For some x” denotes a function that assigns truth to an open sentence exactly if some object satisfies it; “for all x” denotes a function that assigns truth to an open sentence or propositional function exactly if nothing fails to satisfy it.) Richard Montague generalized Frege’s analysis to other connectives.[146] Many contemporary linguists and philosophers of language (myself included) see tense- and case-markers as denoting second level functions. Despite years of tradition to the contrary, the distinction between categorematic and syncategorematic items – lexical and non-lexical items – is plausibly seen as a distinction *within* the class of referring terms. “Or”, “not”, and the like, denote second level functions; “Fred”, “rabbit”, and the like denote individuals or first order functions.

To sum up, a case can be made that so-called syntactical expressions *do* refer. If this is correct (and, as we saw, even if it isn’t correct), then one plainly needs knowledge of semantics to engage in a purely syntax-driven piece of reasoning.

syntactic form and non-syntactic terms

But let us leave aside consideration of terms like “or” and “not”. Even with this waiver, one must *still* know an enormous amount about the

semantics of (iv) to know that it expresses a logical truth. Consider a language L that is just like English except that, in L, “Fred” has the meaning borne by the English word “and”. In L, (iv) is simply meaningless; for it has the same meaning as the English pseudo-sentence:

(ix) “If and is an elephant, then and is either an elephant or and is not elephant”.

Thus (iv) expresses a logical truth only because “Fred” has a quite specific kind of meaning, only because it falls into the semantic category of singular term. Here is another way of looking at it. Let L* be a language just like English except that, in L*, “Fred” has the meaning borne by the English term “rabbit”. In L*, (iv) is meaningless; for it has the same meaning as the English sentence:

(x) “If rabbit is an elephant, then rabbit is either an elephant or rabbit is not elephant”.

(iv) expresses a logical truth in English only because “Fred” has an extremely specific *kind* of meaning: it denotes an individual. If it denoted a second-order function (“and”) or a sortal (“rabbit”), then (iv) wouldn’t be true or false, let alone logically true; it would be ill-formed and meaningless.

The so-called “logical particles” in (iv) are not the only ones that are responsible for its syntactic profile. “Fred” denotes an individual. If it didn’t, then (iv) would be ill-formed, and would thus have an entirely different

syntactic profile from the one it actually has. To recognize that (iv) is logically true, one must know that "Fred" is a certain kind of noun-phrase of some kind; actually, one must know that it is an extremely specific kind of noun-phrase.

Granted, one doesn't have to know *which* exact individual "Fred" refers to; one need only know that both occurrences of "Fred" refer to *somebody* or other. (Actually, one must also know that they *co-refer*: a fact that, as we will see, deepens the connection between semantics and logical form and that further scuttles the attempt to make syntax-driven operations have a foundational cognitive role.) But one must know that "Fred" falls into a highly restrictive semantic category; and there is no way to read this fact off of the morphology of that expression.

A physical object – a deposit of ink, a series of hand-signals – may have one meaning in English, a different meaning in Albanian, and no meaning at all in Greek. Consider the following sentence-token:

(xi) "snow is white".

In English, that token means *snow is white.* Let L be a language in which (xi) means: *if Adam Smith's economic theories were correct, then the democratization of South America would not have been nearly as tortured a process as it actually was.* Considered as an expression of L, (xi) has one syntactic structure; considered as an expression of English, it has a radically different syntactic structure. Given any physical object, no matter what its morphology, it could have any syntactic structure at all. Indeed, given any physical object O, no matter what its morphology, and given any syntactic structure S, there is some language L such that O has S in L. By itself, morphology imposes no constraints at all on syntax; it no more imposes

constraints on syntax than it does on semantics. This is to be expected since syntax is, as we have seen, an aspect of semantics. The only way to have any idea what syntactic form an expression has, one must have a reasonably high-resolution knowledge of its semantics. To know the syntactic form of (xi), one must know that “white” is a predicate, “snow” is a referring term, and so on. And to know that

(xii) “something is white”

can be inferred from (xi), one must know (inter alia) the semantics of “something”. Syntax-drivenness is a kind of semantics-drivenness: so syntax-drivenness is very much *not* innocent. Given a being that lacks a backlog of semantic knowledge, we can no more ask it to engage in syntax-driven operations than we can ask it to write poetry. So it makes dubious sense to regard syntax-driven operations as fundamental.

tokens versus types

Sentence-types are abstract objects, and therefore do not have morphology. Sentence-*tokens* are what have morphology. The sentence-type “I am tired” has no morphology. But any given token of that sentence has morphology.

But even though it is sentence-tokens, not sentence-types, that have morphology, two tokens of a given sentence-type can have absolutely nothing in common at the level of morphology. “I am tired” can be tokened by noises, ink-marks, hand-signals, light-beams, or smoke-signals. No series of hand-movements has any significant similarity to a deposit of ink.[147] Indeed, that sentence-type can be tokened by two deposits of ink that have very little

in common. Consider the various different deposits of ink that token that sentence. Imagine all the different fonts and systems of short-hand that might be involved.

At the same time, two physical objects may be morphologically identical but token entirely different sentences (and thus have entirely different syntactic forms). There is some (possible) language in which an ink-mark morphologically identical with "Bob snores" means *if penguins liked classical music, they would prefer to Bach to Beethoven* and thus has an entirely different syntactic structure from the one it has in English. Given any physical object O, no matter what its shape, and given any syntactic structure S, there is some language L such that, in L, O has S. This means that, by itself, morphology imposes no constraints on syntactic structure.

There is *a* connection between the syntax of an expression and its morphology. But that connection is indirect, and it lies in the *semantics* of the expression involved. Suppose that, when conversing with another English speaker, I make the noise "Bob snores". In that context, that noise has a certain syntax. But it has that syntax only because it has a certain semantics – only because "Bob" is a singular term, "-s" is a tense-marker, "snore" is a one-place predicate, and so on. If "Bob" had not been a singular term, or "snore" had not been a one-place predicate, then that noise would have had an entirely different syntactic form or no syntactic form at all. So *given* a semantics, there is, at least arguably, a connection between syntax and morphology. But absent a semantics, there is no such connection: in that case, syntax and morphology are completely independent of each other. To be responsive to the syntax of an expression, one must be alive to the relevant semantic facts. For reasons we've already discussed, this makes it implausible to suppose that syntax-driven operations have a foundational cognitive role.

indexicality and the dysmorphism between syntactic and logical form

We have seen that the relation between syntax and morphology is, at best, extremely indirect. In this section we will see that, even if that relation were direct – even if it were one of identity -- that would *still* not allow syntax-driven operations to have a foundational cognitive role. In connection with this, we will encounter new reasons to hold that logic-driven operations presuppose semantic, and extra-semantic, knowledge.

Let us start with some paradigms:

(i) Bob looks tired.
(ii) Bob looks ill.
(iii) Therefore somebody looks both tired and ill.

Consider (i). That sentence contains a tense-marker (the "s" on "looks"). Tense-markers are indexicals. On Tuesday, you say "Bob looks tired". You have spoken truly; for at that time, Bob does look tired. On Wednesday, you say "Bob does not look tired". Again, you have spoken truly; for on Wednesday he does not look tired. Suppose that you didn't say anything between those two utterances. So you have said (albeit over a long period of time): "Bob looks tired…Bob does not look tired". You have not contradicted yourself at all. One could know everything there is to know about the *syntax* of your utterances during that period of time, and not have the slightest idea as to whether you had contradicted yourself. Granted, you said "Bob looks tired. Bob does not look tired". But you did not contradict yourself; and if someone were to think otherwise, she would be guilty of a grievous logical fallacy.[148]

Of course, there *are* cases in which an utterance of "Bob looks tired…Bob does not look tired" is self-contradictory. But to know whether it is a self-contradiction, one has to know a great deal about the context in which both sentence-tokens occur. One has to know that they occur within a relevantly short period of time; one must know that the specious present picked out by the tense-marker in the first coincides, at least to a relevantly high degree, with that picked out by the tense-marker in the second. One must also know that the "Bob" referred to in the second is the same as the "Bob" referred to in the first. None of these facts can be read off of morphology; and, more surprisingly, none of these facts can be read off of syntax. One doesn't have enough *linguistic* knowledge to make a determination as to the validity or invalidity of that statement until one knows a fair amount about the contexts in which the expressions in question were tokened. None of this is narrowly syntactic knowledge.[149]

What we've just said about (i)-(iii) is true of all sentences of natural language. Sentences of natural language typically contain tense-markers; so given only a knowledge of the syntactic structures of two sentence-tokens, nothing can be inferred as to the bearing that they have on each other. Even if, wholly on the basis of its morphology, one could know the syntactic structure of a sentence-token (or series thereof), that would *still* not typically suffice for a determination of its validity.

What we've said about tense-markers applies to indexicals that pick out places and individuals. If you are moving about and you say "here is it warm; here it is cold; therefore there is some place where it is both warm and cold", you have made an erroneous inference, *unless* you know that both tokens of "here" pick out the same place. And this knowledge is quite plainly semantic and, indeed, *extra*-semantic: it is, or involves, knowledge of the environment. It is easy to construct analogues of this point that relate to "that", "he", "you"

and most other indexicals.

The context-sensitivity of sentences prevents a mirroring of logical form in syntax. Logical form can be read off of syntax only in cases where the sentences involved are context-insensitive.[150] If our thinking consists in purely syntax-driven inferences, then those inferences will track logical form only if the expressions on which those operations are performed lack the kind of context-sensitivity just discussed. The expressions mediating our thought would therefore have to be stripped of context-sensitive components: they would have to be "eternal" sentences. But it is a point often made that a language lacking context-sensitive components is extremely *inefficient*.[151] Such a language is useless specifically when it comes to information relating to one's own welfare and to knowledge of one's environment: for such information is replete with references to what is here and now; and – valiant efforts notwithstanding – efforts to "eternalize" such information, to reduce indexical information to non-indexical information, simply don't go through.[152] There are many reasons for this. Here is one of them.

Let L be a language like English except that, in L, there are no indexicals that refer to the present time. In L, you have to have one expression for Monday, July 12, 2005, 2:05 p.m., and another word for Monday, July 12, 2005, 2:06 p.m. and so on. It is deeply unclear whether L would have the expressive power of English. Even if you are thirsty on Monday, July 12, 2005, 2:05 p.m., there is a still a difference between thinking *I am thirsty now* and *I am thirsty on Monday, July 12, 2005, 2:05*. p.m..[153]

Now let L* be a language that is like English except that, in L*, *every* context-sensitive of expression of English has been replaced with a context-insensitive one: L* contains only expressions with fixed referents; it thus

comprises only eternal sentences. For extensions of the reasons just given, L* would have dramatically less expressive power than English. In large measure, languages are useful precisely because they contain context-sensitive expressions.

Also, it isn't even clear if languages like L and L* can even exist; the existence of context-insensitive expressions seems to presuppose the existence of context-sensitive ones. English contains context-*in*sensitive terms that refer to specific times and places and individuals. We have expressions like "1492 A.D.", "John Lennon", "the place that is exactly 23 miles west of the western-most part of London", and the like. But suppose English didn't have any context-sensitive terms, and thus didn't comprise sentences like "*now* it is 2005", and "*that* guy is John Lennon". Under that circumstance, it is hard to see how referents could be assigned to expressions like "2005 A.D", "John Lennon", and other context-insensitive terms.[154] Context-sensitivity is a pervasive, and fundamental, feature of language; and context-sensitivity weakens (and arguably severs) the connection between syntax and logical form[155]; and it makes logical form be even more of a semantic matter, as opposed to a narrowly syntactic one, than even we previously said.

In logic books, one encounters paradigms like:

(a) Socrates is a man.
(b) All men are mortal.
(c) Therefore, Socrates is mortal.

In this argument, the tense-marker appears to have little or no role; and, in this case, there is (at least arguably, though not in my view[156]) a straight-

forward mapping of syntax onto logical truth. The problem is that most sentences are not like this. If you pick an utterance of yours at random ("everybody's here", "I don't think there will be a party tomorrow", "you told me that Adam Smith never told his wife he loved her"), it will probably be replete with context-sensitive components; and, for reasons earlier discussed, semantics is needed to bridge the yawning gap between syntactic and logical form. The arguments in logic books are tendentiously chosen to lack such components.

If only for the sake of argument, let us suppose that our thinking does consist in syntax-driven operations on expressions; and let be "Mentalese"[157] be our name for the language to which those expressions belong. Let us evaluate this supposition in light of the points just made about context-sensitivity. If our thinking consists in purely syntax-driven operations on expressions, then *either* those operations will not be logic-tracking, for the reasons just discussed, *or* the expressions on which they are performed must lack the context-sensitivity characteristic of most natural language sentences. So CTM, in its need to have syntax track logic, requires that Mentalese sentences be frozen and context-insensitive – that they lack the context-sensitivity characteristic of their natural language analogues. But if Mentalese consisted entirely of frozen, eternal sentences, then it is hard to see how it could be of much utility to its user: what the organism needs to know is *specifically* indexical; it doesn't need (merely) "eternal" information; it needs egocentric knowledge of the here and now.[158] [159]

To sum up, if Mentalese comprised only context-insensitive sentences, it would be incapable of mediating just the kind of information that the survival of the organism depends on. But the kind of context-sensitivity that enables a language to mediate such information is exactly the kind that blocks a purely

syntactic characterization of logical truth in that language.[160] [161]

syntax as a mirror of semantics

These points about context-sensitivity give us crucial insight into an important matter. Our minds are content-driven (or, if you prefer, semantics-driven). In any case, our thought-processes are consistent with the representational contents of the thoughts they comprise. Suppose I think *Bob is in Greenland and Greenland is a cold place.* In virtue of that fact, I will probably form the belief *Bob is in a cold place*, but not *Jerry is in a cold place.* Any model of cognition must do justice to this fact. CTM (at least Fodor's version of it) accommodates this by taking the following line of thought.

Syntax "mirrors" semantics. Our minds are syntax-driven. So even though our brains are not *directly* sensitive to semantics, they *are* directly sensitive to syntax. (For syntax ultimately reduces to morphology, and there is no difficulty explaining how a brain could be physics-driven.) And, although syntax and semantics are different things, nonetheless a kind of isomorphism can exist between them: semantics can be "coded" in syntax in roughly the way that spatial facts can be coded in arithmetical facts (in triples or four-tuples of numbers). Thus, the operations of a purely syntax-driven entity, like a brain, can be consistent with the *semantics* of the representations on which those operations are performed.[162]

As we've seen, given a language that contains context-sensitive features,

there is no syntactic characterization of semantics, especially not of the semantic relations that matter most to CTM (those relating to proof, entailment and, more generally, relations of bearing). Let L be any such language. Let S be any sentence (or sequence of sentences) belonging to L such that S is logically true. There will be some sentence (or sequence thereof) S* such that S* has exactly the same syntax as S but such that S* is not logically true. Consider the argument expressed by "Socrates is tall; Socrates is wise; therefore somebody is both tall and wise". Because English comprises tense-markers, and other indexicals, it is very easy to construct sentences that are syntactically indistinguishable from the one just given that express *fallacious* arguments:

(*) "He [pointing to John at time t] is tired; he [pointing to Fred at some other time t*] is not tired; therefore somebody is both tired and not tired."

(*) is doubly fallacious: it is nullified by the fact that the times in question don't coincide, and also by the fact that the demonstratives don't co-refer. If thinking consists in syntax-driven operations, and therefore on expressions of some kind[163], those expressions must be replete with context-sensitive components, for the reasons earlier discussed. And this means that, in general, there is absolutely no syntactic characterization of the features of sentences involved in the determination of logical form; no syntactic characterization of entailment, probabilification, proof, and so on. So if we were syntax-tracking machines, we would be illogical except (at most) in the rare and special cases where context was irrelevant. It is probably appropriate

to end this section by quoting Barwise and Perry[164]:

"[T]here is…something profoundly misleading about the traditional concern over entailments between sentences. For one thing, their concern obscures the efficiency of language, and it is this efficiency that is central to meaning. For example the *sentence* SOCRATES IS SPEAKING does not follow from the *sentences* EVERY PHILOSOPHER IS SPEAKING [and] SOCRATES IS A PHILOSOPHER even though this argument as the same "logical form" (on most accounts of logical form) as [ALL MEN ARE MORTAL, SOCRATES IS A MAN, THEREFORE SOCRATES IS MORTAL]. In the first place, there is the matter of tense. At the very least the three sentences would have to be said at more or less the same time for the argument to be valid…A rather startling consequence of this is that there can be no syntactic counterpart, of the kind traditionally sought in proof theory and theories of logical form, to the semantic theory of consequence. For consequence is simply not a relation between purely syntactic elements."

the nature of linguistic meaning

For the sake of argument, suppose that thinking does consist in operations on expressions of some internal language, which we will continue to call "Mentalese". The expressions in question will not be sentence-*types*. Sentence-types are abstract objects. No brain-state can constitute a sentence-*type*; for brain-states are not abstract objects. A brain-state can constitute, at most, a sentence-token. So if thought consists in operations on expressions, it

consists in operations on sentence-*tokens*, not sentence-types. If I am not mistaken, this fact leads to a serious problem for CTM.

Under what circumstances does a physical entity qualify as a token of this, as opposed to that, sentence-type? Consider the *sound* "snow is white". If that sound were made by a volcano 50,000 years ago, no sentence-token would have been produced; nothing linguistic would have been accomplished. Consider the ink-mark "snow is white". If, by some freak accident, a sand-dune, on Neptune a million years go, were formed with that exact shape, no sentence would have been tokened; nothing linguistic would have occurred. So it is never merely in virtue of its physical form that a physical object qualifies as a token of this, that, or any sentence-type. Of course, form plays a part. *Part* of the reason my utterance of "snow is white" tokens this, as opposed to that, sentence is that it has certain acoustical properties. But physical form is not the whole story.

Suppose that Bob, a cognitively normal speaker of English, produces the sound "snow is white". Here we have a genuine sentence-token. How is this case different from the ones just discussed? Bob made those sounds in a certain socio-cultural context, and in consequence of certain linguistic intentions. Bob's noises qualify as a sentence-token not merely in virtue of their morphology, but also in virtue of their being embedded in a certain psychological, historical, and environmental context.[165] Those very sounds could mean anything or nothing. The reason they mean *snow is white* is entirely that they occur in a specific context.

A corollary is that to have any idea what that sound means, you must know something about the context in which it occurs. Absent such knowledge, those sounds are, from your viewpoint, comparable to the sand-dune and the volcano-noise we discussed earlier.[166] Under that

circumstance, those sounds will have no *logical* form at all, at least not from your viewpoint. Responsiveness to logical form presupposes linguistic knowledge; and linguistic knowledge presupposes the contextual knowledge just described. [167] [168]

A genuinely linguistic response to a physical occurrence involves knowledge of a doubly conditional proposition: *if* an occurrence with such and such morphological properties takes place, *then* that occurrence has such and such meaning, *provided* that it occurs in a certain kind of context.

For the sake of argument, suppose that thinking consists in operations on sentence-tokens. For reasons just discussed, if your brain is to respond to the logical forms of those sentence-tokens, then it must your brain must know a fair amount about the facts surrounding the occurrence of those tokens. Such operations presuppose extra-linguistic knowledge, and therewith a wealth of cognitive and representational abilities. So it makes little sense to suppose that such operations could be foundational. [169]

the concept of an algorithm

The idea that syntax-driven operations might be cognitively foundational has been encouraged by a kind of naiveté regarding concepts like "algorithm", "formal inference", "mechanical procedure", and by a subsequent tendency to mistakenly lump these terms together.

There are algorithms for long-division, multiplication, addition, and so on. Let us consider the algorithm for addition (i.e. for the addition of multi-digit numbers). Here is a widely held conception as to the nature of that

algorithm: it is a conception I wish to combat.

Consider the expressions “172” and “35”. There is an algorithm that enables you to add these terms together, and thus to generate a theorem of number theory, *without* knowing anything about the meanings of the expressions involved, and without having prior knowledge of number-theoretic truth. You line up the “172” and “35” in a certain way. You thus form columns of number-expressions. The algorithm tells you to write an expression with a certain shape underneath those columns: what you are supposed to write is a function entirely of the shapes of the expressions in each of those columns. At no point is any judgment or acumen required: you need only be sensitive to the shapes of the expressions involved. If you follow this algorithm – if you write the right shapes in the right places -- you will produce a correct number-theoretic statement. A machine could be designed that wrote the right shapes in the right places, and that therefore executed this algorithm. Therefore a machine could be designed that did addition and, consequently, that generated correct number-theoretic statements.

There is a serious problem with this view. Any expression -- any inscription, any noise or pattern of light -- could mean anything. In English, “172” denotes the number 172. But it could just as well denote Socrates or the number five or nothing at all. Given any entity at all, there is some language L such that, in L, “172” denotes that entity. Since any shape can be associated with any meaning and can thus have any truth-value, it follows that there is no purely morphological characterization of truth, even for lowly

domains like addition. There can be no algorithm that enables you (or any other entity) to decide the truth-value of an inscription *wholly* on the basis of its shape. Truth-value can be decided, if it all, only on the basis of morphology *and* the semantic background.[170]

Suppose that "1" meant *Plato*, and "2" meant *elephant*, but that otherwise our semantic conventions were unchanged. In that case, "172+35=207" would be false or nonsensical. There are infinitely many (readily constructed, if not currently existing) languages such that, in them, "1" means *Plato*, "2" means *elephant*, and so on. The algorithm for addition assigns truth to (e.g.) "172+35=207" only *if* that inscription is part of a very specific semantic system. So that algorithm is given by – is identical with -- a conditional proposition: *if* "1" denotes 1, "two" denotes two, "+" denotes addition, and so on, *then* there is a function assigning truth to certain patterns of those expressions, and falsity to others, on the basis of the morphologies of those patterns. To apply that algorithm is to recognize one of the consequences of that conditional proposition; more exactly, it is to recognize the bearing of that conditional proposition on the question whether a particular inscription[171] (e.g. "172+35=207") is true or (what is closely related) on the question how a particular inscription (e.g. "172+35") must be completed ("=207") if a true expression is to result. So to apply that algorithm involves recognizing the truth of a *multiply* conditional proposition, one that is at least roughly like this: Given that "1" denotes 1, "2" denotes 2, "+" denotes addition, and so on (in other words, given that such and such shapes have such and such meanings); and given also that, relative to that semantic system, there is a function from such and such shapes to such and such truth-values; and given finally that "172+35=207" has such and such shape; *then* it follows that "172+35=207" has a certain truth-value.[172]

So using the algorithm for addition involves, among many other things, one's taking that inscription to be part of a certain semantic system; it thus involves background semantic (and extra-semantic) knowledge, and is thus not purely "mechanical".

As we noted, given any pattern of shapes, there is some language L such that, in L, that pattern is meaningless or false. In light of this, consider an arbitrary act of shape-manipulation. (But focus on the shape-changing itself: leave out any semantic or extra-semantic knowledge that might lie behind it.) For every language L in which that act of shape manipulation constitutes (or generates) a true statement, there are infinitely many in which it constitutes nonsense or falsity. So to compute is not (just) to manipulate shapes. The manipulation of shapes is, at most, a causal consequence of an actual computation – of one's recognizing the truth of a hyper-conditional proposition of the kind earlier described. The manipulation of shapes could almost be described as a kind of epiphenomenon.[173]

These considerations show that, even if there algorithms for number-theory, history, economics, and so on, that would still not mean that number-theoretic, historical, or economic *thought* could be "mechanized". Applying an algorithm is not a purely mechanical procedure; it is a judgment of a certain kind; and it involves, among other things, background semantic knowledge.[174]

"But surely," we are told "an algorithm is a mechanical procedure, and the application of an algorithm is a mechanical process". The statement "the execution of an algorithm is a mechanical process" is ambiguous. There is a *certain* truth in each disambiguation. But, strictly speaking, neither disambiguation is quite accurate.

Remember that an algorithm is given by a conditional proposition: *if*

such and such a semantic system holds, *then* (relative to that semantic system) there is a morphological characterization of truth (i.e. a function from shapes to truth-values). To apply an algorithm, as we saw, is to recognize the bearing of this conditional proposition on the question whether some inscription (or noise….) is true or false. So there are two components to the application of an algorithm. The first component consists in a recognition of the operativeness of some semantic system. The second component involves the existence of patterns of ink (or light or sound…) having the right shapes. The *second* component can be realized mechanically: it can be realized by a being that hasn't even a flicker of awareness of any kind – by windmills, punch-card automata, what have you. But the first component is not in any sense mechanical.

When we say that "computing", and "applying an algorithm", are "mechanical" procedures, that is either false or it is an abbreviated way of saying that the *second* of the two components of applying an algorithm can be realized mechanically (without any sensitivity to semantics, without any judgment).

There is another, slightly different (but related) interpretation of the statement "the execution of an algorithm is a purely mechanical affair". Suppose you want to figure out some sum, say the sum of 172 and 35. There are (at least) two different ways you could go about doing this. You could reflect very hard on the concepts involved (*172*, +, and so on); and, on the basis of a scrutiny of those concepts, you could ascertain that 172 added to 35 must equal 207. Or you could use a procedure that is much more energy-efficient (and intelligence-efficient). Let us discuss this second procedure. First of all, we obviously have semantic conventions regarding these numbers. These conventions are to the effect that "1" denotes one, "+" denotes the operation of addition, and so forth. Let us refer to the set of these

conventions as S. So S is a kind of language. You know that if you write down the S-expressions denoting 172, 35, and so forth, and you make sure that they are lined up in the right way, then there is a rule telling you how to complete that shape – a rule telling you how to produce a shape that, relative to S, constitutes a true statement regarding what happens when 172 is added to 35. If you take advantage of this rule, you will be spared a great deal of thought that you wouldn't otherwise be spared.

This doesn't mean that *no* cognition is involved. Obviously *some* kind of cognitive processing is involved. You use the information given you by sight (or touch or, conceivably, other sensory modalities) to make sure that the ink-marks have the right shape and are lined up the right way. You also call on your memory; you remember what kind of shape you're supposed to write in what place. Also, you know how to interpret, and symbolize, the shapes. You know that "1" denotes 1, and so on. This knowledge is involved in your symbolizing the operation in question (that of adding 172 to 35) in such a way as to render applicable the rule you wish to exploit. (If you don't know how, in S, the number one is supposed to be written down, then the aforementioned rule isn't going to be of much use to you; you won't be able to apply that rule, to exploit it to perform a computation.) So even when you add in the "quick and dirty way" – i.e. even when you line up the "172", with the "35" (in the right way), and so on, and then dutifully crank out an "=207" – cognition is still involved: sense-perceptions are used, memories of semantic conventions are involved. So that act is by no means mechanical. But there is, granted, *a* sense in which that act is much *more* mechanical – much *less* thought-heavy – than the alternative procedure (i.e. the procedure of reflecting deeply on the number-concepts involved).

(To facilitate further discussion, let us systematically use the term "the quick and dirty way" to refer to the kind of operation that we've just been

discussing: the kind where one lines up inscriptions with such and such shapes, in such and such a manner…the kind where *mathematical* reflection is minimized.)

That alternative procedure – the *non*-quick and dirty method -- involves a fair amount of *mathematical* thought. The quick and dirty procedure involves very little mathematical reasoning. The only cognition involved in the quick and dirty procedure is "easy": it involves processing that anyone (who is not severely damaged) can do "automatically": the production of ink-marks having certain shapes, the lining up of those ink-marks in a certain way, the remembering of some fairly simple symbolic conventions, the sense-perceiving and manual production of certain marks. It is wrong, as we've seen, to describe as "purely mechanical" (if by that is meant cognition-free) the quick and dirty method of adding. But it is *not* entirely wrong to describe that process as *automatic.* In some sense of the word "automatically", you can automatically drive a car, turn on a light, understand a sentence of your native tongue. But that doesn't mean that no cognition is involved, that no processing of representations is involved. When you stop at a traffic sign, or understand a sentence of your native tongue, the operations involved are replete with content; they are by no means representationally or cognitively innocent. Perceptions are involved; information is uploaded from those perceptions (that information-uploading is itself, if cognitive scientists are right, a cognition-heavy matter); that information interacts with background knowledge (e.g. *you must stop at red lights or you will be given a ticket or jeopardize your safety…*). These things – stopping at traffic-signals, understanding a salutation in your native tongue – are "automatic", on some interpretation of the word. But they are not cognition-free.

We tend to run together terms like "automatic" and "mechanical". So perhaps in *some*, extremely loose, sense of the word "mechanical", those

sorts of operations are “mechanical”. But they are by no means cognition-free. Far from it. When you add numbers, using the quick and dirty method, your performance is automatic and thus, in a very loose sense, “mechanical”. But it is by no means cognition-free; and it presupposes an enormous background of knowledge, representational abilities, and information-processing power.[175]

Let us sum up. Sometimes when it is said that adding is (or can be) a “mechanical” process, what seems to be meant is that it is cognition-free and representationally innocent. But *that* statement, we have seen, is absolutely false. As we just noted, when a person adds numbers -- even if she uses the quick and dirty method -- various background representations and capacities are mobilized. On the other hand, when a representationally innocent entity generates a pattern of shapes, that *by itself* does not constitute an act of addition; it constitutes only one of two components of such an act. The other component involves the recognition (on somebody’s part) that *relative* to some operative semantic system, the aforementioned pattern of shapes constitutes a true statement. Under no circumstances is a genuine case of adding a representationally, semantically, or cognitively innocent matter; under no circumstances is it “mechanical” in any strict sense.[176]

We should deal with a possible objection to our general line of thought:

Surely there is a sense in a Turing machine computes. And surely such a thing computes by scanning physical objects – by responding in the right way to their morphologies. Surely what we just said about Turing machines is true (mutatis mutandis) of computers generally. Given that there are such things as computers, it is clear that there is something wrong with your arguments.

If by “compute” is meant “simulate logic-driven operations”, then it is indisputable that various artifacts compute. But what we want to know is: Do such artifacts compute in the sense that they *actually* engage in logic-driven operations? *That* is the sense of “compute” that is relevant to discussion of the foundations of mind.

As we’ve seen, to engage in a logic-driven operation, a thing must be sensitive to semantics: it must *already* have a great deal of semantic and, indeed, extra-semantic knowledge. I’ve never heard anyone, whether an advocate of CTM or not, say that Turing machines *do* respond to semantics. Given that the Turing machine isn’t responding to semantics, it follows that it isn’t computing (in the relevant sense): computing involves being sensitive to logical form; and logical form is buried in semantics. [177]

As we have seen:

(*) There is no purely morphological characterization of syntax.
(*) There is no purely syntactic characterization of logical truth.
(*) The only adequate characterizations of logical truth are therefore, at least partially, semantic.
(*) The only way to sensitize a thing to logical form is to sensitize it to semantics.

Since the Turing machine is not semantics-driven, it is not responding to logical form. Therefore it is not computing (in the relevant sense)[178], even though, from some viewpoint, it acts like something that is computing.[179] [180]

References

Barwise, Jon and Perry, John. 1999. *Situations and Attitudes.* Palo Alto: CSLI Press.

Burge, Tyler. 1979. “Individualism and the Mental” in French, Uehling, and Wettstein (eds.) *Midwest Studies in Philosophy* IV, Minneapolis: University of Minnesota Press, pp. 73-121.

Cain, M.J. 2002. *Fodor: Language, Mind, and Philosophy.* Malden, MA: Blackwell.

Carnap, Rudolph. 1947. *Meaning and Necessity* Chicago*:* University of Chicago Press.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax.* Cambridge: M.I.T. Press.
--1975. *Reflections on Language.* Reprinted in *On Language* (1998). New York: St. Martin’s Press.

Coulter, Jeff. 1983. *Rethinking Cognitive Theory.* New York: St. Martin’s Press.

Fodor, Jerry. 1975. *The Language of Thought.* New York: Thomas Y. Crowell

--1981a. *Representations.* Cambridge: The MIT Press.
--1981b. “Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology”. In *Representations.* Cambridge: The MIT Press.
--1987. *Psychosemantics.* Cambridge: The MIT Press.
-- 1990a “Fodor’s Guide to Mental Representation”. In *A Theory of Content and Other Essays.* Cambridge. The M.I.T. Press.
-- 1990b. “A Theory of Content I”. In *A Theory of Content and Other Essays.* Cambridge. The M.I.T. Press.
--1990c. “A Theory of Content II”. In *A Theory of Content and Other Essays.* Cambridge. The M.I.T. Press.
--1990d. “Substitution Arguments and the Individuation of Beliefs”. In *A Theory of Content and Other Essays.* Cambridge. The M.I.T. Press.
--1994a. *The Elm and the Expert: Mentalese and Its Semantics.* Cambridge: M.I.T. Press.
--1994. “Fodor”. In *A Companion to the Philosophy of Mind.*(Samuel Guttenplan ed.). Cambridge: Blackwell.
--1998. *Concepts.* Clarendon Press: Oxford.
-- 2001. *The Mind Doesn’t Work that Way.* Cambridge: M.I.T. Press.

Kaplan, David. 1978. “On the Logic of Demonstratives”. *Journal of Philosophical Logic* 8 (1978): 81-98.
-- 1989. ‘Demonstratives’. In *Themes from Kaplan.* Edited by Joseph Almog, John Perry, and Howard Wettstein. Oxford: Oxford University Press.

McCawley, James D. 1981a. *Everything that Linguists have always wanted to know about Logic.* Chicago: University of Chicago Press.

Putnam, Hilary. 1975. “The Meaning of ‘Meaning’ ” *Philosophical Papers,*

Vol. II : Mind, Language, and Reality, Cambridge: Cambridge University Press.

Pylyshin, Zenon W. 1984. *Computation and Cognition.* Cambridge: The M.I.T. Press.

Quine, Willard van Orman. 1966. “Quantifiers and Propositional Attitudes”. In *The Ways of Paradox.* Cambridge: Harvard University Press.

Russell, Bertrand. 1990a/1905. “On Denoting.” In *The Philosophy of Language.* A.P. Martinich (ed.). Oxford: Oxford University Press.

Wittgenstein, Ludwig. 1958. *The Philosophical Investigations.* Translated by G.E.M. Anscombe. Oxford: Basil Blackwell.
--1983. *Remarks on the Foundations of Mathematics.* Edited by G.H. von Wright, Rush Rhees, and G.E.M. Anscombe. Cambridge: M.I.T. Press.

Can functionalism be articulated without vicious circularity?

Abstract: Many contemporary materialists are functionalists. According to functionalism, x qualifies as an instance of mental T type wholly in virtue of its *functional role.* But the functional role of anything mental involves having other *mental* entities among its causes or effects. Thus, it appears we cannot get a reduction of the mental to the physical along functionalist lines. David Lewis outlined a way out of this problem. His solution is to consider the class of true psychological generalizations *as a whole* and then to Ramsify-out the mental terms: the result is a sentence that contains no constants referring to the mental. I show that that, although Lewis' solution *seems* to work when it is considered at a very abstract level, it does not in fact work: there is no way to generate to relevant Ramsey-sentence without making *ineliminable* use of mental terms. Consequently, Lewis has *not* shown functionalism to be capable of accommodating materialism; and there is *prima facie* reason to believe that it is *not* thus capable.

A doctrine called "functionalism" is widely taken to give the truth about the relation of the mental to the physical. *Prima facie* functionalism is guilty of vicious circularity. David Lewis is widely held to have shown functionalism *not* to be guilty of vicious circularity.[181] In this paper, I show that Lewis has failed to "de-circularize" functionalism and that, in fact, functionalism *is* viciously circular.

Let us begin by putting these points context. First of all, functionalism is a form of *materialism.*[182] Materialism is the doctrine that mental states of affairs are identical with, or realized by, physical states of affairs. There are

excellent reasons to be a materialist. We needn't review these here.

Type-type-materialism is the doctrine the mental *types* (e.g. the type *pain*) are identical with physical types (e.g. the type *c-fiber stimulation*). Type-type materialism is widely held not to be tenable. Many have the intuition that physical states of different types could realize the *same* kind of mental state of affairs.[183] One's brain doesn't have to be composed of this or that material to realize beliefs, regrets, pains, and so on. Surely, we are told[184], it is possible that creatures whose bodies have completely different physical compositions have beliefs, tickles, desires, and so forth. So what *makes* something qualify as a belief that 2+2=4 is not that it is had this or that physical composition: it is not that it is made out if this or that kind of material. It must be something else.

This leaves token-token materialism. Mental types are not identical with physical types. But any *instance* (or 'token') of a mental type is identical with an *instance* of some physical type. Instances of a given mental type can be identical with, or realized by, instances of *different* physical types. So type-type materialism is false. But any instance of a mental type is identical with, or realized by, an instance of *some* physical type. So *materialism* is still correct.

Given that materialism is correct, but that type-type materialism is incorrect, it follows that what makes something x qualify as (e.g.) a belief that snow is white or a pain or a desire for water is not that x is *made* out of this or that type of material: x's *compositional* properties, as we might put it, are not what make it qualify as (e.g.) a desire for water.

What kinds of properties does a physical state of affairs or object have *apart* from its compositional properties? Well, such things also have *causal* properties. And things with very different compositional properties can have,

at least up to a point, the same causal properties. It wouldn't be hard to create something that, to a high degree of approximation, had the same causal properties as the carburetor in your car, but was made out of a completely different kind of material.

This leads us to functionalism: what makes x be (e.g.) a desire for water is not that it is made of this or that kind of tissue, but that it has the right *causal* properties: it is caused by certain things and, in its turn, causes certain other things. x is caused by dehydration and causes you to ingest the contents of glasses of ice-waters: and x is a sensation of thirst *wholly in virtue* of having those cause and effects.[185] The same is true *mutatis mutandis* for every other mental thing. What makes brainstate y be (or realize) a belief that Socrates was wise is that y has certain causes (certain perceptions conjoined with certain cognitive abilities) and, in its turn, it has certain effects (it disposes you to believe certain other things e.g. Socrates was not a fool or Socrates is smarter than Bob, where Bob is some stupid person who you know).

What makes something be a knife is that it has the right causal properties; when it is pressed against cheese or carrots in a certain way, there are certain results. Similarly, what makes something be a belief that 2+2=4 is that it has certain causes and effects. That is the essence of functionalism.

A few more words on functionalism are in order before we critique it. Instances of a given mental type *characteristically* have certain causes and certain effects. Obviously thirst is *characteristically* caused by certain things (e.g. dehydration) and characteristically causes other this (water-seeking behavior or, at any rate, a desire to engage in such behavior). The functionalist says that what *makes* x be a sensation of thirst is that it is caused by/causes such and such, where such and such is a *characteristic* cause/effect of cases of thirst: such and such cannot be an *uncharacteristic* cause or an

uncharacteristic effect or thirst. It is clear why the functionalist must say this. For the functionalist, there is nothing to being a sensation of thirst (or a belief that 2+2=4 or a desire for fame and fortune…) *other* than having the right causes and the right effects. So if x is caused by something that doesn't *char*acteristically lead to a sensation of thirst, or if x doesn't cause anything that characteristically follows from thirst, that doesn't mean that x is a case of thirst that has unusual causes of effects: it means that x just *isn't* a case of thirst.

Clearly functionalism gives a simple and compelling answer to the question how the mental could be identical with, or realized by, the physical. For functionalism, *anything* qualifies as, say, a belief that snow is white or a desire for world-peace long as it has the right causes and effects. Now there is no a priori reason why a brain-state shouldn't cause just those things caused by (say) a belief that snow is white, and shouldn't be caused by just those things that cause such beliefs. (As Hume pointed out, causal relations are not knowable a priori. You don't know a priori that fire burns. So *a priori* anything could cause anything.) So if functionalism is right, then it becomes clear how brain state *could* qualify as a belief that snow is white; for a brain-state could have the right causes and effects. In any case, there is no a priori reason to think otherwise.

But there is a problem. Among the characteristic causes and effects of mental entities are, almost invariably, other mental entities.[186] A belief that snow is white characteristically leads to the *belief* that the color of snow is not black; a desire for something, plus a belief that one can best get that something by doing such and such, characteristically results in an *intention* to do such and such. (The italicized terms denote mental entities.) So *prima facie* functionalism appears *not* to reduce the mental the physical, but only to reduce the mental to the physical *plus* the mental: which is no reduction at all.

David Lewis figured out how to get functionalism to circumvent this problem. Consider all the characteristic causes and effects of (say) the belief that snow is white. Of course, each one of these antecedents corresponds to a proposition of the form *x leads to/results from the belief that snow is white.* Now *conjoin* all the propositions of the form. The result will, of course, be some gigantic conjunctive proposition containing many occurrences of *belief that snow is white.* Let us refer to this conjunction as the L-sentence for the type *belief that snow is white.* For brevity's sake, let us refer to this particular L sentence as 'Bob.'

This brings us to the next step of Lewis's solution. Replace each occurrence of *belief that snow is white* in Bob with a variable, say x1,and then prefix the resulting open sentence with an existential quantifier that binds all occurrences of that variable (in other words, prefix that open sentence with 'for some x1'). What we get is a giant existence-claim that *doesn't* mention the belief that snow is white, but that says everything that (from the functionalist perspective) is important about that belief. Let us refer to this existence as claim as the 'R-sentence for the type *belief that snow is white*'; or, for brevity's sake, simply as 'Sam.'

This brings us to the last step in the argument. Suppose that you generate the R-sentence for *every* mental type and then conjoin all of those sentences. Let L* be this giant conjunction. The result will be something that doesn't mention any mental entities but that (by a functionalist's lights) fully *describes* everything that is important about them. The only constants in L* will be terms referring to things that are physical and non-mental. So Lewis has done a brilliant job of 'de-circularizing' functionalism.

But there is a problem. A precondition for constructing Sam is that you construct Bob, and a precondition for constructing Bob is that you list, or otherwise identify, all the ways one might come to believe that snow is white.

But you couldn't possibly *list* all those ways. For there are infinitely many such ways. There are, for example, infinitely many qualitatively different perceptions, or sets of perceptions, that could lead you to believe that snow is white. And you could, in principle, learn it through some theoretical apparatus; and there are infinitely many different such apparatuses.[187]

So you'd need to cite some kind of *rule* which *generates* the above mentioned list. But I don't see how such a rule could *fail* to contain the term *belief that snow is white* (or some equivalent, e.g. *either a belief that snow is white or is a round square*). For consider all the possible psychological antecedents (or consequents) of that belief. Let C be the class containing all and only these antecedents/consequents. What do all and only the members of C have in common? Well, they all lead to (or follow from) the belief that snow is white. But you can't mention *that* fact when you're trying to construct Bob. More plainly, Bob cannot contain a term like "belief that snow is white" (or any analytically equivalent term). For Bob's whole *raison d'etre* is to show that such terms are *eliminable;* Bob's whole raison d'etre is, in effect, to *not* contain such terms.

Stephen Stich makes essentially this point[188]:

"Typically, if a person believes that everyone aboard the morning flight to Bamako was killed in a crash, and if he comes to believe that the prime minister of Mali was on that flight, those two beliefs together will cause him to believe that the prime minister is dead. Typically, if a person wants to go to New Zealand, and if she comes to believe that she must get a visa in order to make the trip, she will acquire a desire to get a visa. Endlessly many further examples might easily be constructed. But therein lies a problem, for there is no hyperbole in the prospect of *endlessly many* [Stich's emphasis]

further examples. There are literally infinitely many inferential paths leading both to and from every belief…If a person believes [that Niger is west of Mali] and comes to belief that if Niger is west of Mali then Niger has more than two million inhabitants, he will typically come to believe that Niger has more than two million inhabitants. And so on, ad infinitum. Moreover, obviously, this is just the beginning. (Suppose he believes that if Niger is west of Mali, then Niger grows 5 percent more grain, or that if Niger is west of Mali, then Mali is more politically stable, or…[Stich's ellipsis]) If we are to construct our implicit functional definition of 'believes that Niger is west of Mali' from the conjunction of these commonsense platitudes about inference, then the definition will have to be *infinitely long!* [Stich's emphasis.] Now some stout-hearted souls are not intimidated by infinitely long conjunctions, provided they can be recursively summarized….But it is far from clear that [this] class of inferences…*is* recursively characterizable." [189]

What I have been maintaining is, in effect, that the list of inferences – the list of inferences leading both to and from (say) a belief that *Niger is west of Mali* is -- (i) infinite and (ii) *not* capable of any recursive characterization: *except* a characterization that *employs* the term 'belief that Niger is west of Mali' (or some analytically equivalent mental term).

The point is that, unless we help ourselves to the very mental terms we are trying to define – say, *belief that Niger is west of Mali* or *belief that snow is white* – we cannot recursively characterize the causal role of that belief. (And therefore cannot define it at all: for a non-recursive characterization of it would be an infinitely long, and therefore unstateable, conjunction.)

At the same time, if we *do* help ourselves to that mental term, or some analytic equivalent, then our characterization will not be appropriately

eliminative. We will be stuck with an unreduced mental term and will have failed to state the causal roles of instances of mental types wholly in terms of physical entities.

Let us sum up. You can't give an infinitely long *list.* You can, at most, give some *rule* that generates that list. Bob cannot comprise an infinitely long *list*. So it must contain some *rule* which generates such a list. But is very hard to see how such a rule could *fail* to contain the very kinds of terms that Bob is supposed to *not* contain – in this case the term *belief that snow is white* and analytically equivalent terms.

So Lewis' way of making functionalism be non-circular ends up *not* eliminating mental terms in favor of physical ones. The mental terms end up being *ineliminable.* The very sentence by which Lewis 'de-circularizes' functionalism involves the *non*-elimination of the mental in terms of the physical. So, yes, Lewis has de-circularized functionalism: but at the cost of making functionalism *not* effect a reduction of the mental to the physical. [190]

It must be made clear what the structure of our argument is. It is patently obvious that a finite sentence, or finite list of finitely long sentences, can generate an infinite list of sentences or, perhaps, can generate an infinitely long sentence. (I can say: consider the sentence formed by adding "and show is white" infinitely many times to the sentence "grass is green". Here, at least arguably, I have given a finite rule for constructing an infinitely long sentence.) Any axiomatization of a discipline (e.g. Hilbert's axiomatization of geometry) consists of a finite number of sentences/propositions that generate infinitely many such propositions.

The problem with the sentence we have been calling Bob is *not* that it *generates* an infinitely long sentence or that it *generates* infinitely many sentences. That wouldn't be a problem at all; it would be a virtue. The

problem with Bob lies in the following points. Unless Bob contains the very term that he is supposed to eliminated, then Bob is himself infinitely long. And there is no *rule* by which Bob can be constructed *except* for some rule which contains terms which it is Bob's whole purpose to *eliminate.* So *either* Bob is an infinitely long sentence such that there is no *rule* for constructing him. *Or* Bob is a finite sentence which ineliminably contains the terms which it is Bob's purpose to eliminate.

Of course, Bob is concerned only with the type *belief that snow is white*. But what we said about Bob applies is true of all his counter-parts. So suppose that Larry is to the type *belief that grass is green* what Bob is to the type *belief that snow is white.* Everything we said about Bob is true of Larry. Larry will *either* be infinitely long, and there will be no rule for producing him. *Or* Larry will *contain* the very term that he is suppose to eliminate, namely *belief that grass is green.*

Of course, some might say:

I don't see anything wrong with the idea of a sentence S such that S it is infinitely long and such that there is no rule for constructing S. You've argued that Bob is one of those sentences. I don't see what is so wrong with that. Frank Ramsey said that there were infinitely long series of numbers such that there was no rules for generating those series.[191] So why couldn't there be infinitely long sentences that cannot be generated by finite rules?

First of all, what Ramsey said about numbers is highly controversial. Second, even if it weren't controversial, it wouldn't necessarily show anything about propositions or sentences. The idea of a *sentence* or *proposition* that is infinitely long but that cannot be generated by a finite rule

is hard to make sense of. A proposition is the kind of thing that is grasped; there is, I should think, an internal connection between the concept of being a proposition and the concept of being graspable. But something that is infinitely long and cannot be generated by a finite, graspable rule could not itself be grasped.

But, most importantly, even if we do concede that there are infinitely long propositions that cannot be generated by a finite rule – even if we concede that there are ungraspable propositions[192] – Lewis' solution would still fail. Lewis point is this:

We can show that functionalism is not circular. Here how you do it. You do it (at least in part) by means of the following sentence [which we have called Bob]. Here's how you construct Bob; in other words, Bob is the thing you get by following *these* rules…

So Lewis' whole solution goes through only if Bob can be produced or, in any case, identified in some way or other. Lewis hasn't *actually* given us Bob. So Lewis hasn't given us any proposition at all – not one that is finitely long, not one that is infinitely long. He's only given us a recipe for constructing this proposition. But what we have seen that *either* there is no way to apply this recipe at all *or*, if we *can* apply this recipe, the resulting sentence *contains* the very terms it is supposed to eliminate. For the recipe tells us that Bob is formed out infinitely long lists; for Bob is formed by taking *all* the causes/effects of instance of the type *belief that snow is white* – all infinitely many of them – and replacing *each* occurrence with a variable, and so on. But an infinitely long list things cannot be generated, except by rules. And a sentence cannot contain an infinitely long list; it can contain, at

most, a rule for generating that list. So Bob contains rules for *generating* lists of all the causes/effects of instance of the type *belief that snow is white.* Bob cannot itself contain mention of each individual cause/effect. But we've seen that these rules *themselves* ineliminably contain the terms that Bob is supposed to eliminate. So Bob is formed out of a rule which *ineliminably* contains the term *belief that snow is white* (or analytic equivalents). But Bob's whole *raison d'etre* is to *eliminate* such terms. So if Bob *can* be constructed, then he doesn't do his job (which is to eliminate terms like *belief that snow is white*). But to say that Bob cannot be constructed is just another way of saying that Bob doesn't exist. So Lewis' solution is a failure.

I should deal with a possible criticism:

You have said that you *know* of no rule that generates all and only the characteristic causes and effects of instances of the type (say) *belief that snow is white* (apart rules that themselves refer to that type). I admit it isn't *obvious* what an example of such a rule would be. But that doesn't that such a rule doesn't exist.

I think it can be shown that such a rule doesn't exist. Stich has, in effect, already done this for us. As Stich points out, given any proposition, there are "infinitely many inferential paths" leading both to and from that proposition. To put it more formally, given any proposition P1, there are infinitely many sequences of propositions P2….Pn such that, given P2…Pn, P1 can be reasonably inferred[193]. And given P1 there are infinitely many propositions that can be inferred from *it.* Let us focus on sequences of the type P2…Pn. As we just pointed out, the class of such sequences is infinitely large. As a matter of logic, there is no way to say exactly which sequences of

propositions fall into this class *except* in terms of P1: there is, demonstrably, no recursive characterization of the class of all the propositions (or sets thereof) that entail P1, or that it entails, that does not *itself* mention P1. The rules which generates the list in question (let L be that list) must be some rule like: *A sequence S is on L iff S is such that P1 can be reasonably inferred it* (or some equivalent rule, like: *S is on L iff either P can reasonably be inferred from S or S is a square* circle).

These points about propositions can be mapped onto points about beliefs. Let beliefs in the proposition P1 be beliefs of type B1; let beliefs in the proposition P2 be beliefs of type B2; and so on. Since there are infinitely many sequences of propositions that validate (confirm or entail) P1, so there are infinitely many different sets of beliefs that warrant a belief of type B1 and therefore that, in a sufficiently intelligent being, *would* cause a belief of that type. The same point holds mutatis mutandis for beliefs that are caused by a belief of type B1.

Just as there is no way to identity all sequences of propositions of the type P2…Pn (i.e. sequences of propositions that validate P1) *except* in terms of P1, so there wouldn't be any way to identify all the sets of beliefs that would, in a suitably rational mind, lead to a belief of type B1 -- *except* in terms of the type B1. The only way to identify all the sequences like P2…Pn is through some rule that itself mentions P1. Since there are infinitely many sequences of propositions that validate P1, so there are infinitely many sets of beliefs that would legitimate a belief of type B1 and that, in a suitably intelligent being, would lead to beliefs of that type. And just as there is no way to identify all and only sequences like P2…Pn except in terms of P1, so there is no way to identify all and only sequences like B2…Bn, except in terms of B1. So the only to state the possible antecedents of a belief of type B1 would be in terms of some rule that *itself* mentioned B1 and thus didn't *eliminate*

mention of B1.

How to be a non-Atomist about Conception without being a Holist

By a "concept", I mean one's ability to think about a thing. I am able to have thoughts about Socrates; therefore I have a concept of him. I am able to have thoughts about the number four; therefore I have a concept of that entity.

Jerry Fodor argues that, with a few trivial exceptions, our concepts are "atoms". A grasp of any one thing presupposes a grasp of nothing else. I can have a concept of knowledge without having a concept of truth or belief or justification. I can have a concept of causality without have a concept of space or time or persistence. I can have a concept of Socrates without having a concept of corporeality, location in space or time, movement in space, or causality, personal identity, or sentience.

To sum up: For Fodor, one can have a concept of (practically) any one thing without having a concept of anything else. A grasp of (practically) anything presupposes a grasp of nothing. Our concepts are "atoms"; they are indivisible and ultimate; they cannot (typically) be broken down into other concepts.

Fodor provides three ingenious arguments for this astonishing position. The purpose of this paper is examine those arguments. I believe that all three arguments involve non-sequiturs. Where the first two are concerned, the non-sequiturs are, I believe, fairly obvious. But the non-sequitur involved in the third is by no means obvious; and to expose it, we will have to put forth some substantive *positive* theses about the nature of concepts.

§ Let us start with a terminological point. The term "concept" is ambiguous. The word "concept" can refer to some platonic, non-mental

entity, as in: “anything falling under the concept *square* also falls under the concept *closed figure*.” But the word “concept” also has a psychological sense; it can refer to one’s *grasp* of something, as in: “Locke identified concepts with mental images” or “before 1930, nobody had a concept of a quark”.

We need two different words here, if we are to avoid intolerable confusion.[194] To denote concepts in the *non*-mental sense, let us use the word “concept$_O$” (plural: “concepts$_O$”). To denote concepts in the psychological sense, let us simply use the word “concept” (no subscript). [195]

I will talk about “having” concepts and “grasping” concepts$_O$. I *have* a concept (no subscript) of the number two. And I *grasp* the corresponding concept$_O$.

§ Here is a view about concepts that certainly has at least some initial plausibility. It is the view that I hold and that Fodor *rejects.*

Possession of any one concept presupposes possession of others. To grasp any one thing, one must be able to grasp others. One cannot think about, or mentally represent, a given thing unless one think about others. Consider the concept$_O$ of a square. Suppose some creature had no idea what it was for a figure to have four sides, or what it was for an edge to be straight. Would that thing have any idea what it was for something to be a square? Intuitively, the answer seems to be “no”. So to grasp the concept$_O$ *square*, one must grasp the concepts$_O$ of the number four, of straightness, of flatness, and so on.

Of course, what we just said about the concept$_O$ *square* applies to a grasp any one of *these* concepts. Consider the concept$_O$ of the number four. To have a concept of that thing, one presumably has to have a concept of the

numbers three, two and so on. To have concepts of *those* things, one must presumably have concepts of magnitude, order, and who knows what else.

The concept$_O$ of a square might seem like a tendentious one. A square *is* a structure; to be a square just *is* to have certain structural or ordinal features. So given only that a grasp of *that* concept$_O$ presupposes a grasp of others, perhaps no general truths follow regarding the nature of conception.

But what is obviously true of the concept$_O$ *square* is, at least arguably, true of all other concepts$_O$ – even those that *seem* that most devoid of structure. Consider the concept$_O$ of redness. One might naively think that an otherwise conceptless creature could grasp *this* concept$_O$. All the creature needs is the ability to have a certain kind of phenomenal content – it doesn't need an articulated network of background concepts.

But such a view is dubious. An object that really *is* red -- i.e. that has the microphysical properties common to stop-signs, apples, and the like – can be given to a person of good eyesight through very different sorts of phenomenal contents. You have 20/20 vision. You look at the apple from place p1. The phenomenology of your perception of its color is one way. You look at the apple from place p2. Now the phenomenology of your perception of its color is different. This is the case even if you are looking at the apple under optimal conditions – you're in a brightly lit room, your eyes in perfect working order, and so on. No matter how "optimal" or "standard" the conditions of perception are, there will never be a one-one correspondence between things' actual colors and your phenomenology. A given shade of red will given to you through different phenomenal contents.

Any artist knows this. If you want to depict its color as it appears to you from p1, you use one mix of paint. If you want to depict its color as it appears to you from p2, you use a different mix of paint. This is the case, as any artist knows, not just if you are in a room with bad or uneven lighting: it holds even

in circumstances where the lighting conditions are optimal – where they are unnaturally good.

To see something *as* having a certain color isn't just to have an experience with a certain phenomenology. When you just *see*, so to speak, that something is red, a wealth of background information is operative. Phenomenology is certainly part of the story, but not all of it. The phenomenology has to be appropriately contextualized. And this contextualization seems to presuppose some understanding of the conditions of perception; it thus presupposes an articulated picture of the world. So even a concept of redness – surely our best candidate for a "simple" concept, a concept that doesn't presuppose a wealth of other concepts – is available to only to a creature with a formidable representational endowment.

What about *purely* phenomenal concepts? Consider the experience of tasting honey. Consider what it is like to have *that* kind of experience. Perhaps an otherwise conceptless creature could have a concept of what it is like to have *that* kind of experience. So here, it will be thought, we have found a genuinely simple and foundational concept.

But even here there is room for doubt. First of all, there is a difference between, on the one hand, *being able* to have an experience with that sort of phenomenology, and on the other hand, having a *concept* of that sort of experience. Perhaps the ability to *have* that kind of phenomenal content doesn't require any sort of conceptual background. But that doesn't mean that the same is true of the ability to have a *concept* of that kind of experience. Having a *concept* of that kind of experience would seem to involve being able to *recognize* that occurrence of states with that sort of phenomenology and also have thoughts about such states. Even though (perhaps[196]) little or no cognitive wherewithal is needed to *have* the phenomenology, it is hard to believe that the same is true of the ability to *recognize* the phenomenology

for what it is. To recognize is to single out. Singling a thing out seems to involve knowing how that thing is different from other things. So, at least on the face of it, recognizing a thing involves having a concept of how thing differs from other things. So having a *concept* of a certain kind of phenomenology would seem to involve being able to compare it with other things, and thus to presuppose that one has concepts of those other things. So while the ability to *experience* a certain kind of phenomenal content may be conceptually innocent, the same appears not to be true of the ability to have *concept* of that sort of experience.

Let us state what this extremely brief survey has told us. There does not appear to be *any* concept that we can have without having others. There does not appear to be any concept_O that one can grasp without grasping many others.

But there is more to say. Obviously we don't just grasp concepts_O; we also grasp spatiotemporal objects – rocks, trees, and the like. Let us briefly talk about our concepts of these things.

We know about external objects entirely through sense-perception. A sense-perception necessarily presents things *as* having various properties – as having a certain color, shape, location, state of motion. Strictly speaking, what is given to one in sense-perception is never (just) *a thing*: it is a state of affairs. You don't just see *the rock.* You see a state of affairs involving the rock: you see *an* object of a certain shape, color, relative location, and so on. If a thing is to be represented to you as having a certain color, location, state of motion, and so on, it seems to follow that – at some level, in some way – you have cognitive access to these properties: concepts of these properties constitute your representational palette.

So any sense-perception has a great deal of "descriptive structure", and also presupposes some kind of cognitive access to a wealth of properties.

Since sense-perception is essentially involved in our being able to think about external objects, it would seem to follow that our concepts of such objects are replete with conceptual structure.

To sum up, there is at least something to be said for the following view. A concept of a concepts$_O$ always seems to presuppose networks of background concepts. The same is true of our concepts of external objects. Attempts to find genuinely foundational concepts – concepts an otherwise conceptless creature can have – seem to bear no fruit. Our best candidates for that honor (e.g. the concept$_O$ of what it is for something to be red, the concept$_O$ of what it is to have an experience with such and such phenomenology) turn out to bristle with exactly the sort of structure we were hoping they wouldn't have.

So, on the face of it, there seem to be no *atomic* concepts – no concepts that are not to some degree constituted by other concepts. There seems to be no concept C such that a creature could have C without also having various other concepts. There are no foundational concepts. After all, a foundational concept is one such that possession of it doesn't presuppose possession of others.

§ In *Concepts*, Jerry Fodor argues for the exact opposite of this view. He argues that, with some trivial exceptions, all concepts are atoms. For any concept C, possession of C does *not* presuppose possession of other concepts. In principle, one could have a concept of knowledge without a concept of belief or truth (this is one of Fodor's own examples); of justice without having a concept of sentience; of existence in the external world without having a concept of persistence in space and time or of causality; of Tuesday without having a concept of Monday (again, one of Fodor's own examples).

Fodor's overall view is generally rejected. Most hold that atomism is false. But it has not yet been said, to my satisfaction, just where his

arguments for that view go wrong. In this respect, Fodor is a latter day Berkeley. Everybody thinks that Berkeley is wrong: *of course* there are mind-independent objects. But it is not often that one hears it clearly stated just where Berkeley's reasoning goes astray. Similarly, no one has, in my view, pinpointed the non-sequiturs in Fodor's arguments, though many take it for granted that his conclusion is false.

In this paper, I will present Fodor's three arguments, and I will try to show where they go astray. One those arguments involves a fairly blatant non-sequitur. But the other two are quite sturdy. Identifying the fallacies in them involves producing some positive alternative theories about the nature of conception, and is therefore not a zero-sum affair.

§ First of all, Fodor actually uses the term "conceptual atomism" to refer to two distinct (though related) doctrines.

(i) One's *grasp* of a given thing does not typically presuppose a grasp of other things. No concept presupposes possession of any other concept.

(ii) Concepts$_O$ (note the subscript) are atoms. They cannot be broken down into other concepts$_O$. Consequently, there are (with some trivial exceptions) no analytic truths.[197]

As we discussed, it certainly *seems* that, if one grasps the concept$_O$ *knowledge*, one also grasps the concept$_O$ *true belief*. This is exactly what Fodor denies. In his view, one *could* grasp the concept$_O$ *knowledge* without

grasping any other concepts$_O$.

Fodor admits that there are *some* exceptions to this. A grasp of the concept$_O$ *red cow* is not an atom, as it plainly consists of a grasp of the concept$_O$ *red* and also the concept$_O$ *cow*. The same is true of the concept$_O$ *prime number*. If the concept$_O$ one is grasping is *patently* synthesized out of other concepts$_O$, then a grasp of it is not an atom.[198] But with these exceptions, one's grasp of a concept$_O$ is always an atom.

Fodor's denial of (i) is connected with his denial of (ii). Fodor thinks that concepts$_O$ (with the trivial exceptions just given) have no internal structure. The concept$_O$ *knowledge* has no constitutive connections to the concepts$_O$ *belief* or *truth* or *justification.* Of course, if Fodor is right about that, then we can see why a grasp of *knowledge* wouldn't presuppose a grasp of these other concepts$_O$.

Fodor obviously admits that concepts$_O$ typically *appear* to have a great deal of analytic structure. Obviously *knowledge is always true* appears to be analytic; and so do *2 comes before 3* and *causes precede their effects.* He goes to great and ingenious lengths to explain away this appearance of analyticity.[199] We will consider these efforts soon enough.

Fodor grants that there are *some* analytic truths. The concept$_O$ *red cow* is "patently phrasal", as Fodor puts it. It is *obviously* built out of other concepts$_O$ (*red* and *cow*). Any such concept$_O$ will clearly be associated with various analytic truths. (*x is a red cow* implies, and is implied by, the conjunction *x is red and x is a cow*.) Fodor grants the existence of analyticities of that kind. And Fodor grants that *x is a cow* entails *x is a cow.* Notice that those analyticities are *trivial.* They don't expose any hidden depths on the part of the relevant concepts$_O$. In Fodor's view, all analytic truths are in the same category as:

(*) *if x is a red cow, then x is a cow.*

We can be a little more precise about this. (Here we are going beyond what Fodor himself says.) In (*), the non-logical terms occur inessentially. Suppose you replace "cow" with "horse". The result will be a truth. Suppose you replace "red" with "green" or "square": what results will be a truth. The truth of (*) doesn't turn on anything peculiar to the concepts$_O$ *red* and *cow.* Any other concepts could take their place. Since (*) doesn't turn anything peculiar to those concepts$_O$, it doesn't really *tell* you anything about them. It doesn't tell you how *those* concepts$_O$ are different from the concepts$_O$ *horse* and *green.*

Now consider the proposition:

(**) *Any case of knowledge is a case of true-belief (but not vice versa).*

If you replace "knowledge" with "justice", or if you replace "true" with "green", what results is a falsehood (or possibly nonsense). So in (*), "knowledge" occurs *essentially*. The truth of (**) turns on something peculiar to the concepts$_O$ *knowledge* and *belief* and so forth.

Fodor grants analytic status to propositions like (*) – propositions whose truth doesn't turn anything peculiar to the concepts$_O$ occurring therein. Fodor *denies* analytic status to propositions like (**) – propositions whose truth *does* turn on the peculiarities of the concepts$_O$ involved. He appears to grant analytic status to propositions that are *formally* true. But apart from that, he seems to think that there are no, or very few, analytic truths. (Fodor seems to think that *bachelors are unmarried males* is analytically true. But even here he has reservations.)

§ Fodor provides three arguments for atomism. Let us turn to Fodor's first of these. Two thousand years of work have yielded no unexceptionable definitions of 'knowledge', 'justice', or 'number'. In fact, even the most pedestrian concepts$_O$ are indefinable. Consider the concept$_O$ associated with the verb 'to paint.'[200] Fodor goes through a number of reasonable-seeming definitions of this concept$_O$, and shows that they all fail. For example, it might seem that 'to paint' something is just to cover it with paint. But this is not so. If you dip your paint-brush in a paint-bucket, you are not painting your brush, even though you are covering it with paint.

Analogous remarks apply to attempted definitions of practically *any* concept$_O$. Try giving a definition of 'knowledge' or 'justice'. You'll probably find that there is a counter-example to it.

Given the apparent indefinability of most concepts$_O$, Fodor infers that they have *no* structure: they are simples. The idea is that, if a concept$_O$ is not simple, then it is built out of other things, presumably other concepts$_O$. If it is built out of other things, then there is obviously some true proposition saying *how* it is built out of those other things: there is a definition or analysis of that concept$_O$. Since there are no adequate analyses of concepts$_O$, it follows that they are not built out of other concepts$_O$.

On the face of it, this argument concerns *concepts$_O$*, not concepts. It shows that concepts$_O$ – not conceptso – are atoms. But it indirectly supports the idea that concepts$_O$ are atoms. If two concepts$_O$ are not analytically related, then it is hard to see why a grasp of the one would involve a grasp of the other. So if *no* two concepts$_O$ are thus related, then it is hard to see why a grasp of any one concept$_O$ would presuppose a grasp of others. The other side of the coin is that if *any* concept$_O$ is individuated by its analytic relations

to various other concepts$_O$, then presumably one must have some grasp of these analytic relations to grasp any given concept$_O$: so one would have to grasp a multiplicity of concepts$_O$ to grasp a single one of them. So given that Fodor wants to defend the view that concepts (no subscript) are atoms, he is probably right to attack the view that concepts$_O$ (note the subscript) are atoms.

Evaluating Fodor's 1st Argument

This argument involves a non-sequitur. Given his premises, what Fodor should have concluded is that concepts$_O$ have an *infinite* amount of internal structure. Fodor has made a case that it is hard to give conceptually$_O$ necessary and *sufficient* conditions for a thing's falling under a concept$_O$. [201] But for practically any concept$_O$, it is extremely easy to give conceptually$_O$ *necessary* conditions. (In effect, Fodor himself shows this.) Conceptually$_O$ necessary conditions obviously correspond to analyticities. For x to be an act of painting, it is conceptually$_O$ *necessary* that x be an act of putting paint on something. This by itself shows that there is an internal or constitutive relation between 'Bob painted y' and 'Bob put paint on y' and, therefore, that the concept$_O$ 'to paint' has conceptual structure – is not an atom. [202]

Fodor goes through several attempted definitions of the concept$_O$ *paint* (the concept$_O$ corresponding to the verb, not the noun, 'paint'). He points out that, for each one of them, there is something that is *not* an act of painting that falls under that definition. What Fodor does not say is that each of these

attempted definitions gives a *necessary* condition for x's being an act of painting. So the evidence suggests that it is hard to give conceptually$_O$ necessary *and sufficient* conditions for x's falling under the concept$_O$ *paint.* But the evidence also suggests that we can generate *ad infinitum* conceptually$_O$ *necessary* conditions for something's being an act of painting. So given that necessary conditions indicate conceptual structure, it follows that *paint* has an *infinite* amount of conceptual$_O$ structure.

This point applies *mutatis mutandis* to practically any other concept$_O$ one cites. Consider the concept$_O$ *knowledge.* It is easy to generate conceptually *necessary* conditions for x's falling under that concept$_O$. Gettierologists have generated countless such conditions. This is not to mention the ones that have always been known: x must be true and justified and a belief. So it seems possible to generate conceptually$_O$ *necessary* conditions *ad infinitum* for x's being a case of knowledge. If that is right, then the concept$_O$ *knowledge* has an infinite amount of analytic structure, contrary to what Fodor says. If that is wrong, then Fodor is again wrong. Suppose that, after adding some more epicycles, Gettierologists produce a correct analysis of knowledge (x is knowledge iff x is reliably caused, justified, true, well-entrenched….belief). In that case, once again, Fodor is wrong to say that the concept$_O$ *knowledge* has no analytic structure.

§ Let us now turn to Fodor's second argument. Right now, let us focus on one's *grasp* of the concept$_O$ *knowledge.* Let us focus one's concept (no subscript) of that concept$_O$.

Suppose that one's concept of *knowledge* really did implicate concepts of *truth*, *justification*, and *belief* – along with whatever concepts are presupposed by *these* concepts. In that case, we would expect cognitive operations concerning *knowledge* to be slower, requiring of more acumen,

and generally more unwieldy than cognitive operations concerning *justification* or *truth.* But we find no such thing. Thought processes involving one's concept of knowledge are *at least* as speedy and energy-efficient as thought processes involving the concepts that – if the non-atomist view is correct – are constitutive of one's knowledge-concept. Thoughts about knowledge are *at least* as speedy and efficient as thoughts about truth, justification, belief, sentience, and so on.[203]

Of course, we must ask: *why* must it be the case that, if a grasp of *knowledge* implicates a grasp of *truth*, *reliability*, *belief*, and so on, then cognitive operations with *knowledge* would be slower than those with (say) *truth*? Here is Fodor's answer. First of all, when it is said that one's ability to think about knowledge 'implicates' or 'presupposes' an ability to think about (e.g.) truth, what is meant is *not* that there is some merely *causal* connection between the two abilities, but that there is some kind of *metaphysical* relation: *what it is* to be able to think about x as being a case of knowledge *is* (inter alia) to see that x has the properties of being a belief that is non-accidentally true and reliably caused …So for there to be some kind of constitutive relation between one's ability to think about knowledge and one's ability to think about justification or truth is for there to be some *metaphysically* necessary relation between these two capacities.

We know from Hume that there can be no metaphysically necessary relation between x and y unless x *makes up* y or vice versa. (Identity can be seen as a limiting case of this relation.) If x and y are spatio-temporal entities – and surely concepts are such things, for they are realized by configurations of neural matter – then no necessary relation can hold between them unless one of them literally makes up the other, unless, in effect, one of them is a part of the other. But if one's 998-concept, and one's 997-concept… make up one's 999-concept, then presumably in exercising the latter, one is exercising

the former. Of course, to 'exercise' one's 999-concept just is to think about 999. So it seems to follow that, if there are constitutive relations among one's concepts of 999, 998, 997…then in thinking about 999, one is ipso facto thinking about 998, 997, 996…2 and 1. But, as Fodor says, it seems unlikely that I must simultaneously think about 1000 numbers to think about 999. Therefore, my 999-concept is *not* constitutively related to my 998-concept , my 997-concept …or to any other concept$_O$ of mine to which it might initially appear to be constitutively related.

Evaluating Fodor's 2nd Argument

This is a very clever argument. But it covertly assumes the very atomism that it is meant to establish.[204] What the non-atomist maintains is *specifically* that there are no atomic concepts: there are *no* concepts that are not constituted by other concepts. There are no concepts "at the bottom" (for there is no bottom). What the non-atomist maintains is specifically that we mustn't think of the body of our concepts as having a structure like that of a building – where there is a ground floor, a second floor resting on the ground floor, a third floor resting on the first two, and so on. What he maintains is specifically that our concepts support one another (while also – in a way that I hope to describe – being anchored in external reality). The right metaphor is a web, not a skyscraper.

Fodor's argument presupposes that, *if* non-atomism is right, the relation between one's grasp of the concept$_O$ *knowledge* and one's grasp of (say) the concepts$_O$ *truth* or justification is like the relation of the third story of a building to the second story: the one supports the other but not *vice versa.* Fodor seems to think that, if non-atomism is correct, then the relation between

one's grasp of the concept$_O$ and one's grasp of the other is thus asymmetrical. But this is exactly what the non-atomist denies. The non-atomist maintains *specifically* that there are no concepts at the bottom supporting everything else, and also that everything supports everything. He maintains that, while one's grasp of the concept$_O$ *knowledge* is constituted (in part) by a grasp of the concept$_O$ *justification,* the constitution relation goes in the other direction as well. One's grasp of the concept$_O$ supports one's grasp of the other – as different parts of a web support one another. There is nothing at the bottom that supports a grasp of everything else. So, if the non-atomist is right, there is no expectation that operations with the one concept$_O$ should be *more* unwieldy than operations with the other. Since one's various concepts (no subscript) hold one another in place, there is a certain equality among them. None of them is heavier than the other. Fodor's idea is that, if non-atomism were right, then certain concepts would be feather-light, while others would be heavy. There would be those concepts that are fundamental, and thus don't presuppose a baggage of various other concepts: these are the light ones. And there would be those that *do* presuppose this sort of baggage: these are the heavy ones.

But it is just this sort of picture that the non-atomist rejects. What the non-atomist rejects is precisely the idea that there are *foundational* concepts, and a second set of concepts built on *those*, and a third set built on the second, and so on. So what the non-atomist rejects is precisely the view that this kind of asymmetry prevails. For the non-atomist, constitution is typically a mutual relation. Given this, Fodor's argument collapses. One's concept of *knowledge* is constitutively related to one's concept of *justification.* But this doesn't mean that *one* of those concepts must be more basic than the other or, therefore, that one demands more sophistication or energy than the other. Those concepts hold *each other* in place; neither is sitting on top of the other; neither is, in any

absolute sense, simpler or more fundamental than the other. So neither is more wieldy or unwieldy than the other. On the non-atomist picture, constitution is (or can be) a mutual relation. Given only that the one concept is constitutively related to the other, no inferences can be made as to whether operations with the one concept are in any sense more demanding than operations with the other.

Of course, when we are talking about *physical* objects, like statues, the constitution relation is asymmetrical. If the x constitutes y, then y doesn't (typically) constitute x. (In the degenerate case where x and y are *identical*, then perhaps the constitution relation can be thought to go in both directions – but only because, in that case, it doesn't really go in any direction.) So *if* the sense in which one's concept of truth were constitutive of one's concept of knowledge were identical with the sense in which the clay is constitutive of the statue, *then* the constitution relation in the first case would probably be asymmetrical. But when we say that certain concepts are "constitutive" of others, we are probably not talking about the kind of constitutive relation that holds between the clay and the statue. The latter is, at most, a metaphor for the former – and it is (I think we are already seeing) a very misleading metaphor.

§ Fodor's 3rd argument is the most compelling.

For the sake of argument, suppose that some kind *non*-atomism is the case. In that case, a grasp of the concept$_O$ *square* presupposes a grasp of various other concepts$_O$ – *four, straight, planar*, and so on. The idea is that to think of something as a square, you must think of it also as flat, four-sided, and so on. So, for the non-atomist, to think of x as a square just *is* for you to accept various other things about x. Put another way, you belief that x is a

square exactly if you accept that x is flat, four-sided, and so on. The content of a belief supervenes entirely on what you accept in virtue of having that belief – on what you are prepared to infer. Thus, for the non-atomist, beliefs are individuated by their "inferential roles". Let C be some brain-state of yours. For the non-atomist, C qualifies as a belief that x is a square exactly if, in virtue of having C, you are prepared to hold that x has four-sides, that x is not a hexagon, that x is planar, and so forth. So C's content supervenes entirely on C's inferential role. Thus non-atomism presupposes what might be called "inferential role semantics" (IRS).

(For reasons I will give, I myself do *not* think that non-atomism requires that some kind of inferential role semantics be correct. Right now I am presenting Fodor's argument.)

Fodor says that IRS unacceptable for a couple of related reasons. Let b1 be my belief that x is a square and let b2 be Gauss's belief that x is a square. Gauss knew a lot things about squares that I do not. So what he was prepared to accept from the proposition *x is a square* is very different from what I would be prepared to accept from that proposition. So b1 has a very different inferential role from b2. For IRS, this means that, contrary to our hypothesis, b1 and b2 don't have quite the same content – at most, they have *similar* content. But obviously they do have the same content: they are both beliefs to the effect that x is a square.

Let us generalize this point. For an advocate of IRS, differences in inferential role mean differences in content: unless there is identity of inferential role, there is not identity of content. Given this, let B and B* be any two beliefs. Unless B and B* have precisely the same inferential role, they do not have the same content. The trouble is that practically *no* two beliefs have precisely the same inferential role. Let Smith and Brown be two mathematicians who are in exact agreement as to the mathematically

significant properties of the number two. Given *n=2*, Smith will, whereas Brown will not, accept the proposition *n=the age of my son when he learned how to read*; and Brown, whereas Smith will not, accept the proposition *n=the number of pages of my most impressive proof*. So, for IRS, the two beliefs in question *don't* have the same content. But obviously this is false – obviously they have precisely the same content: they are both beliefs to the effect that n=2. For IRS, those two beliefs have, at most, *similar* contents.

In general, for IRS, unless there is *identity* of inferential role there is, at best, *similarity* of content. But no two cognitive states have *identical* inferential roles. So there can be no two cognitive states that have the same content: there can, at most, be similarity of content. So two people never grasp the very *same* proposition: thus there is nothing that two people agree or disagree about, for there is no commonly grasped content. But this is surely false. Surely many different people can grasp the same $concept_O$ and the same proposition.

In connection with this, IRS seems to be committed to an untenable form of holism. Let b1 be an arbitrary belief had by Smith, let p1 be the proposition which is believed.; Let b2 be an arbitrary belief had by Jones; and let p2 be the proposition thereby believed. Any two differences in what Smith and Jones believe will be reflected in a difference in the inferential roles had by b1 and b2. There are two reasons for this. Suppose Smith believes *I am in Paris right now* whereas Jones does not. In that case, Smith will be prepared to infer the conjunction *I am in Paris right now and P* whereas Jones will not be prepared to make that inference. So *any* difference in what is believed can readily be reflected in a difference in inferential role. So if IRS is right , any difference in what two people believe (or in what a given person believes at two different times) amounts to a *total* difference in what is believed and, more generally, what is grasped. So unless two people

are in *total* agreement about everything, then there will be *nothing* that they grasp in common. So IRS makes it totally impossible for two people even to *grasp* the same proposition: there can be no agreement or disagreement about anything.

But this is absurd, as previously noted.

There is another related problem with IRS – sometimes it is known as the "error-objection" to IRS. If the inferential role of a concept *determines* its content, then inferences involving that concept cannot be *wrong*: since they *define* the concept, and give it its content, they cannot be wrong *relative* to the content of that concept. Imagine the following. There is some function F such that you have the following beliefs about it. F(4)=6, F(8)=10, and so on: for each x less than a thousand, you believe that F(x)=x+2. But you also believe that F(1000)=1004, and that F(1006)=1010, and so on. If IRS is right, then the fact that you take F(1000) to equal 1004 means that the function you were grasping *all along* was something other than F(x)=x+2. You were grasping some other function – perhaps *for all x less than a thousand, F(x)=x+2, but for all x greater than a 1000, F(x)=x+4*. The point is: if IRS is right, you cannot *wrongly* apply a function that you grasp; for the beliefs you have *determine* which function you are grasping. So IRS makes it impossible for to *wrongly* apply a concept: wrong applications of a given concept ends up being a correct application of some other concept. If you decide that F(1000)=1004, you are not wrongly apply F(x)=x+2: you are rightly apply some "bent" cousin of that function. If IRS is right, then given any concepto C, an exactly similar argument shows that wrong applications of C collapse into correct applications of some other concepto C*. (If, with Frege, we think of concepts as being identical with functions – functions from individuals to truth-values – then it is very clear how generalize the argument just made.)

But plainly there is such a thing as *wrongly* applying a function. There is

such a thing as wrongly adding two numbers: it is not as though would-be wrong applications of the addition-function collapse into correct applications of the addition* function, where addition* is some bent version of addition. So IRS is wrong.

Evaluating Fodor's 3rd Argument

I think that Fodor's 3rd argument collapses when we apply an analogue of Kripke's distinction between reference-fixing and meaning-giving.

Let us start with the basics. When we say that somebody has a "concept" of some thing x, what we do we mean? We mean, I think, that that person can think about x – that he can entertain thoughts (of some kind or other) about that thing. To have a concept of Socrates is to be able to have thoughts about Socrates – thoughts like *Socrates was bald.* If we say that Sam has a concept of the number two, we are saying, nearly enough, that Sam can think true or false things about that number. It seems self-contradictory to say: "Sam has a concept of the number two. But he just can't grasp truths of falsehoods concerning that number."

Being able to think about a thing seems to involve being able to *single it out*. To be able to think about the number two, you need to be able to cognitively lock onto it, to the exclusion of all other things.

Consider your concept of the number two. It seems inconceivable (to me, anyway) that somebody would have a concept of that number but *not* know that two is one more than one and one less than three. It seems inconceivable that a person would be capable of grasping the proposition *Fred has two apples* without also grasping the proposition *Fred has more than apple* and *Fred less than three apples*, and so on.

This suggests that certain pieces of knowledge are *constitutive* of one's

concept of the number two. One couldn't possibly believe n=2 if one didn't believe *n=1+1* and *n=3-1*. This suggests that for one to believe *n=2* just *is* to believe *n=1+1*, and so forth.

Fodor denies this because he thinks it leads to an untenable holism. But it is obviously something we *don't* want to give up lightly – and I am now trying to show that it does not lead to an untenable holism.

But there is an incoherence in what we've just said. One cannot have true beliefs *about* x unless one *already* has a concept of x. You cannot believe that two is greater than one or less than three unless you *already* have a concept of the number.

So we have to find a way to accommodate two opposing sets of intuitions. First, it seems inconceivable that one would be able to grasp propositions like *n=2* and *John has two apples* unless one could grasp related propositions (e.g. *n is more than one*, *John has more than one apple*). This suggest that having a concept of the number two involves having knowledge of certain rudimentary facts about that number. And this, in its turn, suggests that one's concept of the number is *constituted* (at least in part) by knowledge of such truths.

At the same time, knowing a true proposition *about* the number two involves *already* having a grasp of that number.

There is an easy way out of all this. The appearance of contradiction is generated by a false premise. Consider the proposition

(*) *somebody x was uniquely a great philosopher to die of hemlock poisoning.*

(*) *describes* Socrates. But to grasp (*) it isn't necessary to *already* have a concept of a Socrates.

Let me a little more precise about what the term "describe" means in this context. (*) is an existential generalization, not a singular proposition. But a

generalization is always made true by some *singular* fact. It isn't possible just for *somebody* or other to cough. Some *specific* person must do so. It must be *Smith* who coughs – or Jones or Brown. So the proposition *somebody coughs* is "made true" by the singular fact that Smith coughs. That particular proposition is made true by many different singular facts – by Brown's coughing, Jones's coughing, and so no. Given a proposition of the form *for some x, x has phi*, let us say that O is a "truth-maker" of that proposition iff *O has phi* is true. So Brown is a truth-maker of the proposition *for some x, x coughs* because *Brown coughs is true.*

An existential generalizations of the form *something x uniquely has phi* necessarily has at most one truth-maker. Let us say that a proposition having that form "describes" O iff *O uniquely has phi* is true. Clearly (*) "describes" Socrates (in our technical sense of the word): Socrates is its unique truth-maker.

Now if you grasp a proposition that describes some thing x, then you are in a position to have further thoughts about x – you are in a position to (truly or falsely) attribute properties to x. Suppose Smith thinks to himself: *somebody x was uniquely a great philosopher to die of hemlock poisoning.* Smith can then think: *and that guy probably had a most unpleasant death*. In fact, for any concept_{O} phi that Smith grasps, Smith can think *that guy had phi*. Because Smith knows the truth of (*), Smith is now in a position to attribute any property he wishes to the truth-maker of (*). In a word, Smith can think anything he wants about Socrates.

You are talking with your friend Smith, whom you know to be a reliable information. Smith says to you: "Somebody just discovered a cure for AIDS." You don't know *who* discovered the cure. But you are still able to say things like "well, *that* person will be famous", "*that* person has done humankind an immeasurable service", "is *that* person working for a government agency or

for a private organization?", "I bet *that* guy has a thing or two to say about current government policy regarding stem-cell research". Basically, given knowledge of a proposition that *describes* (in our technical sense) someone or something x, you are able to think things *about* x.

What we just said about Socrates and the (future) discoverer of a cure for AIDS can be applied to anything. Consider the number two. Suppose you know the truth of the following proposition:

(**) There is some number n such that n=1+1.

If you know (**), then you are in a position think things about *that number.* You can think "I wonder if Smith has *that number* of cars" and "*that number* is greater than one" and "any closed figured has more than *that number* of sides (if its sides are straight)."

Knowledge of a proposition that *describes* a thing apparently suffices for ability to have thoughts about that thing – so knowledge of such a proposition seems to qualify as a *concept* of that thing. And knowledge of such a proposition does not itself presuppose a concept of that thing.

Given all of this, here is what I would propose. To have a concept of a thing is to know some proposition that describes x (in our technical sense of "describes"). Now many will respond by saying that this view has been straightforwardly refuted by Kripke, Putnam, and others. Such a response, I will argue, involves a number of confusions – mainly, a confusion of linguistic meaning with cognitive content. But before I deal with that set of issues, allow me to develop our theory, and also to show why, if correct, it falsifies Fodor's slippery-slope argument.

Suppose you know the truth of (**). First of all, as we've seen, it

immediately follows that you are able to have thoughts about the thing which (**) describes. (**) describes some number n. If you know (**), then you are able to think *n is larger than one* and *Smith has n cars*, and so forth. Second, if you know (**), then you *automatically* know a certain fundamental truths about n – you know that n=1+1, that n is greater than zero, and so on.

Now we can close the argument. Suppose that, as we are maintaining, your knowledge of (**) is your *concept* of the number two. Let C be your belief in that proposition. In that case, it doesn't matter what else you come to believe (truly or falsely) about the number two. Obviously *C* is not, on that account, going to undergo any change; so your concept of that number isn't going to change. It is true, of course, that the inferential role of that concept will be different: now, given the proposition *Fred has two cars*, you will be prepared to infer that *Bob and Fred have the same number of cars*, whereas, previously, you were not prepared to draw that inference. But none of that is going to change your *concept* of the number two in the slightest. After all, that *concept* is identical with your belief in (**) (or some other proposition that describes two), and that belief isn't going to be changed with changes in what you believe *about* that number.

This analysis has another virtue – it sidesteps the "error-objection" to IRS. On our view, not *every* one of the beliefs you have about the number two is definitive of your concept of that number. Only your belief in (**) is thus definitive. But once you have a belief in (**), you can have any number of wrong beliefs about that number. On our view, to think that the number two has phi is to think something like the following: *for some number n, n=1+1, and n has phi.* Now suppose that the number two does *not* have phi. On our view, you are still having a wrong thought *about* the number two. Suppose Brown has three cars. Given this, let phi be the property of being equinumerous with the number of cars Brown has. On our view, to think *2=the*

number of cars Brown has is to think something like:

(***) *for some number n, n=1+1, and Brown has n cars.*

(***) is a wrong belief *about the number two.* Your wrong belief about two doesn't collapse into a correct belief about three. What you've done is to attribute to Brown the property of having n cars, where n=1+1. So what you've done is to *wrongly* think that Brown has *two* cars. So you've had a wrong belief about the number two: your mis-application of the concept *two* didn't collapse into a correct application of some other concept.

Our analysis has one virtue. The objects of thought are not objects but *propositions.* You don't think *Socrates.* You think *Socrates is bald* or *Socrates was wise.* It isn't possible to think a *thing* – only a truth or falsity about a thing. You don't think *two* – you think *two is even* or *1+1=2.* In any case, this seems plausible –luminaries ranging from Kant to Frege to Brandom have held it – and it is worth seeing how far we can go with it.

A concept is just *is* the ability to grasp a thing. But, as we've just noted, grasping a thing consists in grasping a proposition *about* that thing. So it is hard to see how one could exercise a concept of a thing without thinking a truth or falsity *about* that thing. Our analysis is obviously consistent with this: to have a concept of the number just *is* to know a certain kind of proposition about it.

The other side of the coin is this. Suppose you say that having a concept of (say) the number two does *not* consist in knowing some proposition. In that case, it seems to follow that you could *exercise* your concept of that number – you could *think* about the number two – without thinking anything *about it.* Permit me to clarify the logic involved. To have a concept of (say) the number two is to have a grasp of it. But what is grasped is never a thing *simpliciter*, but

only some proposition concerning that thing. So a concept of a thing must be a concept of a proposition *about* that thing. Otherwise we are saying that one can grasp a thing *without* grasping any proposition concerning it: we are saying that one can *just* think *two*, as opposed to *1+1=2* or *two is even.* But, on the face of it, it very much seems that whatever is thought – whatever is grasped – is some truth or falsity, and never just a bare object. In any case, our analysis is on the right side of the view – which seems reasonable to many, though it will not seem so to many others – that it is whole propositions that are thought, and never just bare things.

So far we have (tendentiously) been speaking about numbers. Let us apply our analysis to less sterile domains, beginning with a story. You don't know who invented bifocals. (It was Benjamin Franklin, as it happens.) Somebody whom you know to be a reliable source of information says to you: "Some one person invented bifocals." On this basis, you are able to have thoughts *about* that person. You can say things like: "Well, I bet *that* person was very proud of his invention" and you can ask things like "did *that* person have poor eyesight? Did that person personally benefit from his discovery?" What is going here seems to be this. You know some true claim of the form: something x uniquely has phi. In this case, the proposition in question is:

(IBF) some man uniquely invented bifocals.

That proposition describes (in our technical sense) some man – namely, Benjamin Franklin. Because you know this proposition, you are able to affirm things, and ask things, and so forth, *about* that person. So your knowledge of (IBF) gives you the ability to think about that person – it is a *concept* of that person.

When you *use* your knowledge of (IBF) to attribute some property phi to that

person, you are really thinking some existence claim: *some man uniquely invented bifocals, and that person had phi*. So when you use that concept to attribute cleverness to that person, you do so by way of the thought: *some man uniquely invented bifocals, and that person was clever.*

Here we must be very explicit, and very cautious, if we are to stay on the right side of some important semantic and modal points. Your concept of the person who invented bifocals is constituted by your knowledge of proposition that describes that person. You grasp him *through* the proposition *unique inventor of bifocals*. But this doesn't corrupt your modal intuitions in the slightest. You know that *that person*, whoever it was, didn't *necessarily* invent bifocals – that he might have died at the age of three or might have decided to choose a path in life that didn't involve inventing bifocals.

Suppose I say to you: "some guy stole my wallet" (i.e. some man x uniquely stole my wallet). For analogues of the reasons just given, you are able to have thoughts *about* that person, whoever it is. You can think "he must be a person of low morality" or "he probably had a bad childhood" or "he is probably a drug addict". So even though your concept of this person is constituted by your knowledge of the existence claim *somebody x uniquely stole __'s wallet*, you still know that that person, whoever it was, didn't *have* to become a thief – that he might have taken a different path in life. He might have become a monk or a university professor. So the fact that, in at least some cases, we know things through existence claims which describe them does not force us to have the wrong intuitions about modality or analyticity.[205]

Let us take stock. First, as we saw, your knowledge of (IBF) constitutes a concept of Franklin. Second, that concept is obviously not an "atom". It presupposes a grasp of other concepts – the concept of bifocals, of being an inventor, of being a male, and so forth. (Of course, possession of any given one of these conceptso presupposes possession of others: consider what you

must *already* grasp to grasp the concepto of bifocals or of being an inventor.) The third point is the most important. Now that you know (IBF), and thus have a concept of a certain person, it doesn't matter what you *henceforth* think about that person: it won't alter your concept of him. Suppose you learn that was also witty: that won't change the concept in question: the concept is constituted by your knowledge of (IBF) – it isn't going to turn into a new concept, or otherwise change, when you learn new things about him. When you learn that that person was one of the Founding Fathers, your *concept* of that person isn't going to change. It is true that, under that circumstance, you will accept some inference that you didn't previously accept: you will accept the inference from *x invented bifocals* to *x was a Founding Father.* But this fact doesn't mean that your concept of that person suddenly changed: that concept remains what it always has been – a knowledge of an existence claim. By the same token, you can come to believe any number of *wrong* things about that person: you can come to believe that he was over seven feet tall or that he invented the airplane. Those beliefs will be wrong beliefs *about* Franklin, not about some other person or thing (and not about *no* thing either). Remember what we said a moment ago: when you exploit your knowledge of (IBF) to attribute a property phi to the person it describes, you are really thinking an existence claim: *some man uniquely invented bifocals, and that person had phi*. So suppose you use your knowledge of (IBF) attribute the property of being over sever ft tall to that person. You do this by thinking the following existence claim:

(IBF7) *some man uniquely invented bifocals, and that person was over seven ft tall.*

In thinking (IBF7), you are attributing a property to the inventor of bifocals

that that person didn't actually have. You are having a *wrong* thought about that person. But this doesn't threaten the integrity of your concept of that person. Roughly speaking, the concept is the left half of (IBF7); the false belief is the right half. You can accumulate any number of false beliefs about the inventor of bifocals, and that won't corrupt your concept of the man: put loosely, the concept stays one side of the comma, and the subsequently acquired wrong beliefs remain safely sequestered on the other side of it.

So here we have a concept of somebody that is *not* an atom and that is not vulnerable to any of the arguments that Fodor (or anyone else) directs against IRS.

Now our analysis of conception does seem to involve *a* kind of holism. As we said, to grasp IBF, you must obviously grasp various concepts$_O$. First, you must grasp the concepts$_O$ immediately implicated in that proposition – *bifocals*, *inventor*, and so on. But, as we noted, a grasp of any one these concepts$_O$ presupposes a grasp of others. To grasp the concept$_O$ of an inventor, you must (I would think) grasp the concepts$_O$ of causality, agency, intentionality (in both senses of the word). To grasp the concept$_O$ of bifocals, you must grasp the concept$_O$ of eyewear, of bad vision, and therefore of vision itself; and a grasp of the concept$_O$ of vision involves the concepts$_O$ of space, time, causality (again). So the ability to grasp (IBF) – the ability to have that sort of concept of Franklin – is by no means innocent: it presupposes a background of conceptual weaponry. But it does not follow, and it is not the case, that our analysis of conception is guilty of an untenable form of *belief* holism.

§ If our analysis is to be at all compelling, we must deal with a major issue. We have argued that to be able to think about an object O just *is* to know some true proposition of the form *something x uniquely has phi*, where phi is

a property had by O. We have argued, in effect, that to have a concept of O is to have knowledge of a *uniquely individuating description* of O.

But in light of some points made by Kripke, it is generally taken for granted that this is simply false. Thanks to Kripke, the following is generally taken for granted: one can think about O *without* having such knowledge; conception is *not* descriptive in nature; in many cases, the descriptive information one associates with an object is incomplete, and thus applies to many things besides that object, or doesn't even apply to that object at all. So it would seem that our "descriptivist" analysis of conception is on the wrong side of some very powerful and widely accepted views.

The viewpoint just described has its basis in two points made by Kripke. The first is a narrowly semantic point – a point about the literal meaning of proper names (expressions like "Socrates", "Benjamin Franklin", and the like). The other point is epistemological.

§ Let us start with the semantic point. It is a datum that "Socrates" refers to Socrates. Kripke (1972) made a strong case that proper names are not synonymous with definite descriptions. Contrary to what was previously thought, "Socrates" is not synonymous with "the great philosopher who drank hemlock" (or any other definite description).

This strongly suggests that "Socrates" does *not* refer to Socrates by way of any description, and that it refers to him *directly.* Thus "Socrates" is *directly referential*: it is a mere label for a person, not a description that singles out that person. From this, it seems to follow that to *understand* "Socrates" is to know what it refers to: nothing more, nothing less. To understand that expression is to associate it with the right *individual*, not some description or Fregean sense.

Any English speaker can *refer* to Socrates by means of the word

"Socrates". Presumably, if a person can understand sentence like "Socrates was wise", then that person is able to grasp what is meant by that sentence. And that, in its turn, would involve grasping the things meant by the components of that sentence – in particular, by "Socrates". For reasons stated a moment ago, it seems that the thing meant by "Socrates" is simply Socrates – it isn't a description or Fregean sense that singles out Socrates.

This brings us the Kripke's epistemological point. Kripke argued[206] that people don't typically know "uniquely individuating descriptions" of the objects to which they refer and about which they think. Any English speaker can *understand* the sentence "Socrates is wise". But if asked to produce a description that uniquely applied to Socrates, nine out of ten people couldn't do so. They would be able to produce *some* description, but not that one applied to Socrates to the exclusion of everyone else, and perhaps not even one that applied to him at all. They might say "he was the author of the *Metaphysics*" (false) or "he was a famous philosopher of antiquity" (incomplete, applies to people besides Socrates)

This seems to show that knowledge of a description that singles out x is *not* necessary for thinking about x. Smith may be able to think about Socrates and refer to him; but the only *description* Smith can associate with Socrates is some faulty or overly inclusive one ("author of the Metaphysics", "philosopher of yesteryear", "philosopher who died of unnatural causes").

§ Let me begin by identifying my position – this will involve stating my biases. First, I myself *agree* wholeheartedly with Kripke's semantic point. "Socrates" is *not* synonymous with "the great philosopher who drank hemlock" or any other definite description. Kripke's arguments for this are extremely compelling. In connection with this, I also *agree* that "Socrates" is directly referential: at the level of semantics, "Socrates" is simply a label for

Socrates. So the proposition meant, at the level of literal meaning, by "Socrates was bald" is simply *Socrates was bald*: it isn't *somebody x was uniquely a great philosopher who drank hemlock, and x was bald* or anything of the sort.

Here is where I *disagree* with the story told in the last few pages. First of all, literal meaning is not epistemology. We must be very careful to distinguish what is *literally* meant by an expression, on the one hand, from what it communicates to us. I am not alluding to the kinds of non-literal meanings discussed by Grice, but to something much more radical. There is a difference between hearing the *sound* "Socrates was wise", on the one hand, and knowing what it means, on the other. One has to compute the meaning of the sounds; and this inevitably involves a great deal of background knowledge. The background information *through* which computes literal meaning is to be distinguished from literal meaning itself. Much of what a sentence-utterance communicates to us is connected with that background information. Kripke is (I believe) exactly right about what is literally meant by "Socrates": that expression is not a definite description. But given only that fact, nothing follows regarding the information *through which* one assigns the right literal meaning to "Socrates was bald", and other sentences containing "Socrates". And, I will argue, the information *through* which one assigns the right literal meaning includes knowledge of the right kind of uniquely individuating description. That description will seldom, if ever, be a historically momentous one, like "the great philosopher who drank hemlock"; it will typically be rooted in the peculiarities of the context in which one first encounters the relevant referring term. But, I will argue, knowledge of a uniquely individuating description is *always* involved in knowledge of what "Socrates" means and what it refers to. Kripke's argument to the contrary has weight only because it focuses on the wrong uniquely individuating

descriptions.

§ The best way to begin is with Kripke's own positive account of reference and conception. In his view, one is able to refer to Socrates, and to cognitively access the man, *not* by virtue of knowing some individuating description, but by virtue of having a certain *causal* connection to the man. [207]

The main idea is that a *causal*, rather than a descriptive, connection to x is what underwrites the ability to refer to, and think about, x. And this idea seems to have a prototype in the fact that to *sense-perceive* an object x is not (merely) to have a mental image that "fits" x, but is to have a mental image that has the right *causal* connection to x. (If you have a mental image with the right phenomenology, but the wrong cause, it isn't a perception of the dog in front of you – it is a hallucination or, perhaps, a perception of some other dog.)

§ What is the theory of reference that Kripke outlines? [208] [209] Let's say I don't know of any uniquely individuating description of Socrates. Still, I can refer to and think about the man. How? Somebody encounters Socrates for the first time. (Presumably this is when Socrates is born, but that doesn't matter in this context.) That person dubs his new acquaintance "Socrates". [210] Socrates goes about, introduces himself. Some people have the good fortune to *personally* see Socrates and be told that he is called "Socrates".

Those people obviously can refer to Socrates. And they can *transmit* that ability to others. They can say: *I knew this fellow named "Socrates". He was the most obstreperous person I ever met.*…Having heard (or even just overheard) this, the auditor of that statement can *himself* refer to, and think,

about Socrates. This kind of transmission can go on ad infinitum. 2,500 years after Socrates' death, we can refer to him, thanks to a long chain of such transmissions.

§ Let us now critique this picture, beginning with a purely terminological point. In this context, I will talk a great deal about "descriptions". Sometimes we use that word to denote certain kinds of linguistic entities, e.g. "the man who stole my wallet", "a great general of antiquity". In this context, I will use that expression to refer to the information *expressed* by such linguistic entities; so I will use the term "description" to refer to the concepts expressed by things like "the greatest French soldier" and the like.

Now to matters of substance. Suppose that your companion points someone out to you and says: "that guy is Fred". As a result, you learn who "Fred" names. Here it *seems* as though you learn who "Fred" names *without* knowing any true proposition of the form: *something x uniquely has phi and "Fred" names x.*

But that is not the case. Whenever an object is successfully indicated to you, you *do* have knowledge of a uniquely individuating description of it. If your friend's ostensive definition of "Fred" is to work, you must be able to single out the thing ostended. He cannot just be *a* guy. He must be some specific guy. And he is – the right kind of uniquely individuating information is easy to come by. He is the unique guy standing *over there*, next to *that* water fountain. Or he is the unique object of your attention at *this* moment in time. Or he is the unique guy now running away from you with your wallet.

The uniquely individuating description through which you lock onto Fred may be, and probably is, highly context-specific -- replete with the kinds of things that, in language, are expressed by demonstratives (*that* guy over *there*…). But here what is important is that the lock-on involves knowledge

of *some* description that singles Fred out: the exact nature of this description is irrelevant.

Let us sum up this leg of the argument. Your friend says "that [pointing to someone] is Fred". If you are to assign "Fred" to the right person, you must have some way of cognitively singling that person out. Doing this involves knowing some description that Fred alone satisfies. That description may be, and probably is, highly context-specific and thus of no historical moment (*the unique guy who just bumped into my car and then drove away without giving me his insurance information…*) But that is irrelevant in this context – what is relevant is you know of *some* uniquely individuating description.

There is more to say. An ostensive definition typically involves an indexical –something that, in English, usually involves a "this" or a "that". Now an indexical always picks out the thing satisfying a certain description. "That man" picks out the thing satisfying the description *contextually salient male* or perhaps *ostended male.* "Today" picks out the thing satisfying the description *day of utterance.* "I" picks out the thing satisfying the description *speaker of utterance.* The same holds (mutatis mutandis) for any other indexical. So order for an indexical to work – in order for it to direct the auditor to the right thing -- the thing thereby referred to has to satisfy some existence claim: it must be a thing x such that x uniquely satisfies a certain description. (Let IDL be an arbitrary indexical-token.) And the auditor assigns the *right* referent to IDL only in so far as he knows some true claim of the form *there is some x such that x uniquely satisfies description D, and IDL refers to x.* So even *ostension,* and thus ostensive definitions, operate through the medium of uniquely individuating descriptions: and assigning the right referent to the ostension involves knowing the right such description.

§ Let us now reconsider Kripke's story in light of these platitudes.

Imagine that there people who personally *saw* Socrates, and were told: *that* guy is named "Socrates". (To simplify discussion, let us naively suppose that the word they heard really was "Socrates".) Each of these people knows some true proposition of the form: *somebody x uniquely has phi and "Socrates" names x.* For the reasons considered a moment ago, each one of these privileged few learns the truth of a proposition like: "Socrates" denotes *the unique person **over there**, next to the hearth, talking about the One, wearing a purple toga…*

Some of these people might actually meet Socrates. In that case, once again, one has a uniquely individuating description. Each of these people learns the semantics of "Socrates" through a description like: "Socrates" denotes *the person who **I** shook hands with at time t, place p, who irritated **me** by contesting my definition of piety…* (The description has reference-fixing, not meaning giving, role.)

Suppose that Larry is one of the people who where given an *ostensive* definition of "Socrates". So Larry *saw* Socrates, possibly met him, and was told: ***that*** guy, over there…is named "Socrates".

Larry's uniquely individuating description almost certainly has an *indexical* component (the indexicals are in boldface). Larry is told that "Socrates" refers to ***that*** guy. Larry *sees* (or otherwise sense-perceives) Socrates: So Socrates is given to Larry through information like (speaking from Larry's perspective): *the guy who **I**'m talking to right **now*** or *the guy ten feet from **me**...*

But that doesn't in any way mean that Larry doesn't have a uniquely individuating description of Socrates. He *does* have such a description. Many uniquely individuating descriptions contain an indexical component.[211]

Let us continue the story. Larry outlives Socrates. Talking to his buddy Fred, Larry says (in Ancient Greek): *I knew this guy named "Socrates"; that*

guy did such and such…

As a result, *Fred* has a uniquely individuating description of "Socrates", namely: *the guy to whom Larry was referring on occasion C (at place p, time t)…*

Here a couple of points are in order. First of all, there is no circularity in this account. Larry had a uniquely individuating description of Socrates. That is beyond doubt. Now Larry's description was not circular: it did not have the form: "Socrates" names the guy called "Socrates". Larry's uniquely individuating description (uniquely individuating description) was something like: "Socrates" refers to *that* guy – the one next to the hearth in the purple toga…

Now Fred's uniquely individuating description of Socrates isn't circular either. Fred's uniquely individuating description is something like: "Socrates" denotes *the guy to whom Larry was referring on such and such occasion.*…So Fred's concept of Socrates consists in Fred's knowledge of the proposition: *there is some x such that, on such and such occasion, Larry was referring to x, and "Socrates" names x.*) Fred's uniquely individuating description is parasitic on Larry's. But Larry's description isn't circular, and neither is Fred's.

Another point must be made. Larry could introduce Socrates through a highly impoverished description: *But even so, Larry has a uniquely individuating description of Socrates.* This point is crucial and, so far as I know, has always been overlooked.

Suppose Larry says: *once I met this chap named "Socrates". He was a real jerk. I wish I'd never met him. I'm sorry I brought him up. Let's not talk about him anymore.*

It might seem at first that, under the circumstances, Fred doesn't have a uniquely individuating description of Socrates – that he has only an overly

inclusive description: *some jerk who Larry met.*

But that is not so. Fred *does* have a uniquely individuating description : *the guy who Larry was referring to at place P, time T*…For there was *one* guy who to whom Larry was referring at that exact instance – at the specific time when Fred first heard the name "Socrates". So the description through which the semantics of "Socrates" is given to Fred (in the reference-fixing, not the meaning-giving, sense) *is* uniquely individuating and accurate, namely: *the guy who Larry was referring to at place p, time t.…*

Further, there is no way that the ability to refer to Socrates could be transferred from Larry to Fred (or from *any* person to any other) *except* by way of a uniquely individuating description (I mean a description that *in fact* singles out the relevant object). Transference of that power, from Larry to Fred, involves, at the very least, Larry's saying something about Socrates to Fred, or in Fred's presence. And no matter how cursory that something is, Fred *ipso facto* has a uniquely individuating description of Socrates. Suppose that, in Fred's presence, Larry utters

(*) "that Socrates was a real nuisance. I wish he'd just stayed in Athens."

Given that (*) is Fred's only source of information about Socrates, it might *seem* as though Fred doesn't even come close to having knowledge of a uniquely individuating description of that philosopher: all Fred has is the incomplete, and possibly false, description "a real nuisance". But this is not so. Fred knows that, on such and such occasion, there was some x such that Larry was referring to x and also that "Socrates" names x. So given only the meager bit of information embodied in (*), Fred *does* know some true proposition of the form *something x uniquely has phi and "Socrates" names x.* In this case, the proposition in question is: *there is some x such that, on*

such and such occasion, Larry was referring to x, and "Socrates" names x. [212]

There are different ways one can upload a concept of a thing. The uploading may be accomplished through an ostensive definition ("*that* guy is named 'Socrates'") or through a descriptive definition (" 'Socrates' is the name of the great philosopher who died of hemlock poisoning". It may accomplished through some threadbare bit of (possibly erroneous) reportage, like (*). But no matter which of these ways is operative, one's concept

The other side of the coin is that if Fred has *no* uniquely individuating description of Socrates, then he *isn't* connected to Socrates in the kind of way he'd have to be to refer to him or have (de re) thoughts about him. The amount of information needed to acquire knowledge of a uniquely individuating description of Socrates is minimal. One need only hear a passing reference to the man. (Of course, seeing Socrates will do.) For then one has the uniquely individuating description: *the guy referred to on occasion C* (or *the guy I saw next to the Oracle of Delphi, gesticulating wildly while talking about the nature of piety…*) So if one doesn't have a uniquely individuating description of Socrates, then one doesn't even have the basis for the minimal connection needed to refer to Socrates or to think about him (in a de re way).

The main point is this. Larry transfers certain abilities to Fred: the ability to use "Socrates" to refer to Socrates, and the ability to think about Socrates in a de re way. Larry's transferring these abilities to Fred involves Fred's acquiring a uniquely individuating description of Socrates.

Of course, we can imagine Fred transmitting the powers in question to some third party – say, William. By the logic we just discussed, William's

acquiring these powers involves his (William's) acquiring a uniquely individuating description of Socrates. And so on, until the chain we've described reaches you, or your grandchildren.

It must be pointed out that the only *correct* uniquely individuating description that Fred has may be one like: *the guy to whom who Larry referred on such and such occasion.*…Beyond that, Fred's beliefs about Socrates may be quite spurious. Fred may believe that Socrates was a person of low morality, and that he wrote the *Posterior Analytics*, and so on. By the same logic, the only uniquely individuating description one of us – twenty-five hundred years later – has of Socrates may be one like: *the guy who Frank was referring to that time when.*…Beyond that, one's beliefs about Socrates may be spurious.

So one *does* have to have a uniquely individuating description of Socrates to think about, or refer to, the man. But that uniquely individuating description may be highly specific to one's own circumstances, and may thus be of no interest to historians or anyone else who was interested in Socrates' biography. Russell – *rightly*, as I think we are seeing – thought that to think about, or refer to, Socrates one had to have a uniquely individuating description of the man.[213] But, judging by his examples[214], he *wrongly* took the uniquely individuating descriptions in question to be historically momentous ones: *the guy who invented the theory of forms and drank hemlock.* He was wrong about the nature of those uniquely individuating description, but right, apparently, that one must have *some* uniquely individuating description.[215]

So be able to think about and refer to Einstein, it is necessary to have *some* uniquely individuating description of the man. But that uniquely individuating description may be of no historical moment; it may be, and

probably is, something like: *the guy who so and so was talking about on such and such occasion.*…That is why one can be so wrong about what Einstein did – about whether he invented the atom bomb, for example – *and still refer to and think about Einstein.*

Notice, also, that this account involves a *causal* connection between Socrates and between people today who think about and refer to him. In other words, the account we've given *is* a causal account, even though it is also a descriptivist account. Larry *sees* Socrates: obviously there is a causal connection there. And that is how Larry gets *his* uniquely individuating description of Socrates. Fred sees or hears Larry refer to Socrates on some particular occasion: obviously there is a causal connection there. And so on, until this chain reaches you. So you *are* causally connected to Socrates: and it *is* this causal connection that enables you to refer to and think about Socrates. But this causal connection is embedded in various uniquely individuating descriptions (or perhaps it would be better to say that these uniquely individuating descriptions are embedded in it).

§ Clearly not just *any* causal connection between a person and a thing enables that person to refer to that thing. The causal connection operates *by way of* some conception. I am directly causally connected to many individual atoms (which bounce off many surfaces). But given some one of those atoms – call it "Alpha" – can I refer to Alpha? Is my being hit by Alpha *enough* for me to refer to it? It seems not. I have to have some way of *singling out* Alpha in thought. That way of singling out Alpha may – indeed probably will –exploit my causal connection to Alpha. But it is not the causal connection *by itself* that enables me to think about or refer to Alpha. It is some conception which (probably) *uses* that causal connection.

§ These points map onto – and derive considerable support from – the case of sense-perception. To see the sun, I must indeed be causally connected to the sun. But *how* does that causal connection give me cognitive (visual) access to the sun? It cognitively connects to the sun *through the indexicals in a certain piece of existential-descriptive information.* You don't just see *the sun*; you see a total situation involving the sun. Your visual perception "says": *that* [indicating the sun] *is an object with such and such properties (it is luminescent…)*…Or: *over there* [indicating a certain place, above the horizon perhaps] *is a luminescent object.*… The "that" and the "there" obviously involve a causal connection. And they connect you, cognitively, to the object. But they operate in the context of an informational-tableau that gives you a uniquely individuating description of the sun. The uniquely individuating description will be something like: *the unique luminescent object* ***there…***Or: *the unique luminescent celestial body* ***I*** *am* ***now*** *attending to…*

All perception is "predicational", as Barry Stroud (2001) says. You don't just see *Fido.* You see *a* thing *with such and such properties*….The content of your perception is not given by a noun, but by an existence-claim[216]: *over there, there is a four-legged furry…*[217] It would be impossible to see something as being property-less. Any situation visually represented is represented as having properties. So the content of your perception is not just some *thing* – it is not just *the sun* or *Fido* or *Dad.* What is represented to you is *a* situation having various properties. When you do what we call "seeing the sun", what you see doesn't correspond to the noun-phrase "the sun"; it corresponds more closely to a whole sentence: *there is a bright orb over yonder…*[218]

Of course, you can see the sun only if you are causally connected to it. But

the causal connection operates *within* a uniquely individuating description : it is a *part* of a uniquely individuating description. The content of your perception of the sun is at least in part, given by existential information: *there is a bright orb over* ***there*** *(in* ***that*** *region)…*Said information has an indexical-component (or, more accurately, a linguistic representation of it would have such a component), corresponding to the "over there, in *that* region". The causal connection between yourself and the sun is represented in the presence of this indexical component. So the causal theory of perception ("I see the sun because I am causally connected to it") and what we might call the "descriptivist" theory of perception ("I see the sun because my visual state encodes a description that *applies* to the sun") are by no means incompatible: they are both true. And the same is true of the relationship between the causal theory of reference ("I can refer to Socrates because I have a certain causal connection to the man") and the descriptivist theory of reference ("I can refer to Socrates because I know of a description that singles him out"). They are not incompatible; they supplement each other; they are two parts of a single true story. [219]

§ Now we can return from the steamy jungles of epistemology to the cool waters of semantics. Given what Kripke argued (not to mention Putnam, Soames, and others), it seems pretty likely that *Hesperus has phi* has for its literal meaning the proposition: *Hesperus has phi.* But nobody will deny that *Hesperus has phi* and *Phosphorous has phi* have (or may have[220]) very different cognitive values. Some have reacted to this latter fact by trying to come up with a new and better form of descriptivism. But given what we said a moment ago, I don't think such a drastic measure need be taken.

We saw that (notwithstanding what is, or seems to be, alleged in *Naming*

and Necessity) in order to know what is meant by “Socrates”, one *does* have to know some existence of the claim of the form: something x uniquely has phi, and “Socrates” refers to x. Now when you hear a sentence-token, you don’t *automatically* know its meaning. In principle, one can hear the sound without knowing the meaning. (This would happen if you didn’t speak the relevant language or if you had sustained brain-damage of a certain kind.) You must make your way from the sound to the meaning. You can’t do that in a vacuum; you have to exploit your background knowledge. As we saw, if you know what (to switch examples) “Hesperus” refers to, it is by way of some existence claim, which might be something like:

(#H) *yesterday there was a unique star x such that I was gazing at x and x alone remained in the morning sky after all the others had disappeared; moreover “Hesperus” names x.*

If you are to understand an utterance that you hear of *Hesperus has phi*, you have to hook up that token of “Hesperus” with the right referent. And to do that, you have to exploit the existence claim that embodies your knowledge of what it refers to. So to assign the right referent to “Hesperus”, you have to work through some existence claim like (#): that existence claim (or some other) is your only lead. So even though what *Hesperus has phi* literally means is simply *Venus has phi*, nonetheless you can assign it the right meaning only via (#). So what “Hesperus is lovely” ends up imparting to *you* is some descriptively rich existential proposition:

(#HL) *yesterday there was a unique star x such that I was gazing at x and x alone remained in the morning sky after all the other stars had disappeared; moreover “Hesperus” names x*; and *x is lovely.*

For exactly similar reasons, what "Phosphorous is lovely" conveys to you will be something like:

(#PL) *yesterday there was a unique star x such that I was gazing at x and x alone appeared in the evening sky before all the other stars; moreover "Phosphorous" names x*; and *x is lovely.*

So even though "Phosphorous is lovely" has for its literal meaning some descriptively impoverished singular proposition (*Venus is lovely*), nonetheless what it conveys to you – thanks to the descriptive information through which you *assign* it the right meaning – is some descriptively rich existence claim. It is clear why, given these principles, "Hesperus=Phosphorous" will communicate a complicated existence claim, even though what it literally means is some bare identity.

§ This analysis is on the right side of Kripke's powerful modal and semantic points. Consider the sentence:

(*) "Hesperus is the last star to disappear from the morning sky."

Kripke makes the (correct) point that (*) is not *analytic*. "Hesperus is *not* the last star to disappear from the morning sky" is not self-contradictory.

Our analysis is consistent with this. Let us suppose, once again, that the proposition through which you assign "Hesperus" to Hesperus (Venus) is this:

(#H) *yesterday there was a unique star x such that I was gazing at x and x*

alone remained in the morning sky after all the others had disappeared; moreover "Hesperus" names x.

Notice that "Hesperus" gets *narrow-scope* with respect to the descriptive information. Nothing – no descriptive information, no Fregean sense – comes between *Hesperus names* and the thing named. To use Kripke's important distinction, the descriptive information has a "reference-fixing" but not a "meaning-giving" role. So you know that "Hesperus" is *just* a name, and *not* a description, and that all it does is to label a certain thing. So on our analysis (*) is *not* analytic – at the level of literal meaning, it encodes a proposition of the form *O the last star to disappear from the morning sky.* It does not encode a proposition of the form *the last star to disappear from the morning sky is the last star to disappear from the morning sky.*

Further, on our analysis, you *know* that (*) is not analytic: you have the right intuition about this. You assign "Hesperus" to Hesperus by way of your knowledge of (#H). In that proposition, the descriptive information is given "wide-scope". So you know that no descriptive information comes between "Hesperus" and the thing it labels; you know that "Hesperus" is *just* a label, and not a description. So you know that (*) is not analytic. So you have the right intuitions where the modal status of the *sentence* is concerned.

In connection with this, our analysis is consistent with the fact that you *also* have the right intuitions about the modal status of the *proposition* meant by (*) and also about the modal properties of the thing named by "Hesperus". Suppose that the (#H) is the proposition through which you affix "Hesperus" to right entity. In knowing (#H), you know that "Hesperus" names the celestial body which *in fact* disappears from the morning sky after any other. But you will believe that the thing which "Hesperus" names *had* to have that property. In virtue of knowing (#H), you know that something x *in fact* is a

celestial body that disappears after all others from the morning sky. Your knowledge of (#H) is not knowledge to the effect that something *had* to have that property. That knowledge is consistent with the position that the thing which *in fact* had that property might not have had it at all. For any object x having that property, your knowledge of (#H) is consistent with your believing the truth of the proposition *x might not have been a last celestial body to disappear from the morning sky.* Put in positive terms, for any object x, your believing (#H) is consistent with your believing the proposition *x might not have been a last celestial body to disappear from the morning sky.*

Another example might be appropriate. Suppose you believe the proposition

(**) "Benjamin Franklin" names the unique person x who *in fact* invented bifocals.

There is no object x such that, in virtue of believing (**) to be true, you believe that x *had* to invent bifocals. Put in positive terms, for any object x, your believing (**) is consistent with your believing the proposition *x did not have to invent bifocals – x might have done something else with his life.* So our descriptivist account of conception does not saddle anyone with the wrong modal intuitions about propositions or objects.

On our descriptivist account, you do indeed think of Hesperus (Venus) by way of some descriptive proposition. But you still have the right modal intuitions. This is because your concept of Venus consists in knowledge of a proposition of the form *something x* in fact *has phi*: it does not consist in knowledge of a proposition of the form *something x* necessarily *has phi.*

One other point should be made explicit. On our account, it is necessary that you think of Hesperus by way of *some* claim of the form *something x*

uniquely has phi. But there is no *one* proposition of that form such that you must think of Hesperus by way of that particular proposition. On the contrary, it is essential to our analysis that it be possible to think of Hesperus by way of *different* propositions of that form. That feature of our analysis is essential to our way of solving Frege's puzzle. So our account does not by any means make it necessary that you think of Hesperus *as* the last celestial body to disappear from the morning sky. It allows that you can think of it through infinitely many different descriptions.

That said, it might seem that there is *still* a sense in which our analysis is not modally innocuous. On our account, you think of an object O s *by way of* your knowledge of proposition of the form *something x uniquely has phi*, where phi is a property had by O. So, on our account, you think of Hesperus through *some* such proposition. Of course, our account allows that there are several such propositions. But, for the sake of argument, suppose that in your case, there is only one such proposition; and suppose it is:

(****) *something x is uniquely being a last celestial body to disappear from the morning sky.*

So on our account, when you think *Hesperus has phi*, you are really thinking:

(*****) *something x in fact has the property of uniquely being a last celestial body to disappear from the morning sky*, and x has phi.

So what you think when you think *Hesperus is lovely* is:

something x in fact has the property of uniquely being a last celestial body to disappear from the morning sky, and x is lovely.

A special case of this principle is that, when you think

(A) *Hesperus is the last a last celestial body to disappear from the morning sky*,

what you are thinking is:

(B) *something x is uniquely being a last celestial body to disappear from the morning sky*, and *something y is uniquely being a last celestial body to disappear from the morning sky*, and x=y.

Here is the apparent problem for our analysis. The proposition:

(A_N) Necessarily, *Hesperus is uniquely a last celestial body to disappear from the morning sky*,

is simply false, and nobody thinks otherwise. But what about (B)? At first glance, (B) looks like the kind of proposition that might be necessary. If that is the case, then our analysis is false.

The truth is that our analysis does not put us on the wrong side of any modal facts.

Let us consider the operator *necessarily.* This operator can combine in different ways with (B): more exactly, it can be given varying degrees of scope.

(B1) **Necessarily**: *something x is uniquely a last celestial body to disappear*

from the morning sky, and x is identical with a *unique a last celestial body to disappear from the morning sky.*

(B2) *Something x is uniquely a last celestial body to disappear from the morning sky*, and **necessarily**: x is identical with a *unique a last celestial body to disappear from the morning sky.*

(B1) and (B2) are simply false; and nobody would ever think otherwise. So in the relevant respects, they just like (A_N). So our analysis does *not* turn true modal claims into false ones, or vice versa.

Nonetheless, it might seem that our analysis has a peculiar consequence. We have said that you think of Hesperus through either (****) or through some other proposition of the same form. So it seems that, on our analysis, you *must* think of Hesperus a certain way: you *must* think of it as a last celestial body to disappear from the morning sky or a first celestial body to appear in the evening sky: in any case, you must think of it through *some* uniquely individuating description. But this seems wrong: for there doesn't seem to be any that anyone *has* to think of Hesperus. For reasons we've seen, it seems that cognitively "locking onto" a thing O involves knowing some proposition of the form *something x uniquyely has phi*, where O has phi. But after that lock-on has been established, I don't *have* to think about O by way of that proposition. I can drop the proposition. Suppose that I lock onto Sam by having a visual perception of him. In that case, as we saw, I *do* lock onto Sam by way of some propostoin of the form of *something x uniquyely has phi*. In this case, that proposition might be *somebody x is uniquely a person standing over there, at this time…*or perhaps *somebody x is uniquely a person whom I am focusing on at this time…* So forging a concept of a Sam involves

knowing some such propositoi. But once that concept has been forged, I can, and probably will, forget that proposition. Suppose Sam becomes a life-long acquaintance of mine. I will almost certainly *not* remember the proposition through which I initially locked onto him. So it *seems* that, after a certain point, the uniquely individudating proposition can simply be dropped: after which point, one has a "pure", non-descriptive, object-tracking concept. So, after a certain point, one does *not* have to think of Sam (or anything else) by way of some descriptive proposition. If that is right, then obviously our anslysis is false. After all, our analysis nearly enough just *is* the claim that conception is knowledge of some such description.

Here is what I will argue. Any given descriptive proposition can be dropped, and probably will be. But at no point does concpetoin become "pure". What happens is simply that one's knowledge of the old proposition is replaced with knowledge of a new (but related) proposition. At any given point in the career of one's concept of a thing, one's ability to think about that thing consists in one' s knowing *some* such proposition. But there is typically not some *one* such proposition such that, any any point in the career of one's concept of a thing, one must think about that thing through that specific proposition. Let us now turn to this matter.

§ One can easily forget the information *through* which one first acquires a conception of an object. Indeed, this is what happens most of the time. I meet Sam in 1985. At that point, I am told: "that [pointing to some guy on the jungle gyms, wearing a hat with a propeller on it…] is Sam". So it is indeed through some kind of existence claim that I hook up the name to the right person: *there is some guy x who was just pointed out to me, and whom I am now looking it, who is over there on the jungle gyms, wearing a hat with a*

propeller on it…and "Sam" names x. But it would be very easy for me to forget that "start-up" existence claim. In fact, with very few exceptions, I *never* remember those start-up claims; they drop out.

This might be taken to show that, at some point, I acquire a "pure" conception of Sam: a conception that is no longer paved with knowledge of existential-descriptive information. So the idea would be this. Perhaps knowledge of a uniquely individuating description was needed to *initiate* the concept. But after that, it isn't needed anymore.

I think this is a big mistake. The "start-up" existential information certainly does drop out. I may not remember, even sub-consciously, the circumstances under which I met Sam. But this absolutely does *not* mean that, at any point, I have some kind of "pure" conception of him, i.e. one that is not paved with existential information. Indeed, if such "pure" conceptions were possible, it is hard to see how it could be possible for Hesperus/Phosphorous cases to arise. But, I fully grant, the point in question definitely requires that we supplement our analysis.

Suppose my first encounter with Sam is through existential information like: *there is guy wearing a green shirt and a bowler hat, whom I am now (on occasion t) meeting and…*and, consequently, that such information is, at least initially, implicated in my interpretetings of "Sam".

I meet Sam again, four days later. And I do so with knowledge: I *recognize* him. What is happening here? My recognizing him obviously doesn't consist in my becoming aware of some bare identity; it doesn't consist in its being the case, that for some object O, I suddenly become aware that *O is identical with O.* No doubt we do recognize the truth of such bare identities. But that is not what *recognition* consists in. Recognition is not so informationally impoverished.

The story I would tell is this. I realize that there was some chap whom I met

a few days earlier, on occasion t, who had such and such features (e.g. he was wearing a bowler hat)…and I also realize that, right now, on occasion t*, there is some chap, who has thus and such properties (e.g. he is wearing a baseball cap)…and I also realize that there is some *one* individual who uniquely satisfies both of these existential propositions.

Let E be the first existence claim, and let E* be the second.

We see that recognizing a thing consists in seeing that some one thing uniquely satisfies each of *two* existence claims. This means that E -- the "start-up" existential information -- can be dropped. My *recognizing* Sam involves my seeing that some one entity satisfies both E and E*. Having made this recognition, I am at liberty to drop E. My awareness of Sam will be sustained, for the time being, by my knowledge of E*. And, in its turn, E* may be dropped and replaced by a third existence claim.

So what sustains a conception of an entity is knowledge of a series of interlocking existence claims. Any given one of these claims may be dropped, and probably will be. But at any juncture in the career of one's having a conception of a thing, that conception involves knowledge of *an* existence claim. And this point, as we've seen, enables us to account for the cognitive difference between "Hesperus is lovely" and "Phosphorous is lovely".[221]

The nature of emotions in relation to the central question of meta-ethics

I. Let us refer to sentences like ‘Stalin was evil’ and ‘murder is wrong’ as *valuative* sentences. By a ‘valuative’ sentence I mean one that expresses a moral attitude or sentiment. Different ways of analyzing such sentences have been proposed. In this paper I wish to show that two such analyses that are thought to be incompatible may in fact be compatible. More specifically, I will try to show that a doctrine known as ‘emotivism’ *may* be compatible with – may indeed be a certain kind of -- a doctrine known as ‘moral realism’ (or ‘realism’ for short). Emotivism, I will argue, is a kind of realism *provided that* a certain plausible thesis concerning the nature of emotions is accurate. I will provide some case for that thesis, though I will certainly not be able to establish it with anything approaching certainty.

According to emotivism, statements like ‘Stalin was evil’ express emotions or emotionally laden attitudes towards objects or states of affairs. If you say ‘Stalin was evil’ you are expressing some kind of negative emotion or attitude towards Stalin.[222]

According to realism, sentences like ‘Stalin was evil’ ascribe properties to objects. The sentence ‘Stalin was evil’ ascribes a property to Stalin, just like the sentence ‘Stalin was short’ and ‘Stalin was born in Tblisi.’ So ‘Stalin was evil’ is as much capable of truth or falsity as ‘2+2=4’ and ‘snow is green’. [223]

It is generally taken for granted that emotivism and realism are incompatible.[224] This viewpoint is correct, I will show, *only if* emotions are not beliefs. If emotions *are* beliefs, then emotivism is a form of realism.

Or so I will argue.

For the record, I myself think that emotions *are* beliefs and that, consequently, emotivism is a form of realism. I also think that this position is in reasonably good accord with our pre-theoretical intuitions, even though it might *seem* at first to conflict with them. In any case, the view that emotions are beliefs or judgments is fairly widely held by philosophers of mind (though it is probably not a majority position), and its bearing on ethics should be investigated.

A couple of disclaimers are in order. My argument works only if certain controversial, but highly plausible, assumptions are granted. These assumptions are not innocuous, and I will not be able to provide compelling arguments for them in such a short paper. But I will *outline* arguments for them, if only to show that they deserve some credence and that my ultimate destination – proving that emotivism may well be a form of realism – is one that *might* be legitimately arrived at. When I make use of one of these controversial assumptions, I will try to indicate that fact.

One other qualification is in order. The words ‘good’ and ‘bad’ have meanings that relate to morality as well as meanings that have nothing to do with morality. Sentences like ‘this is a good machine gun’ and ‘Bob is a good assassin’ are obviously not moral statements. ‘Bob is a good assassin’ means ‘Bob is an effective assassin’; the word ‘good’ is being used in an instrumental, not a moral, sense. And, of course, if you say ‘this is a bad atomic bomb’, you are using the word ‘bad’ in an instrumental, not a moral, sense. The words ‘good’ and ‘bad’ also have aesthetic meanings, of course. The aesthetic meanings of these words may be more closely connected to their moral meanings than are the instrumental meanings of these terms. But the aesthetic meanings of ‘good’ and ‘bad’ are still to be distinguished from their moral meanings. When you say ‘that was a good concert’ or ‘Bach was

a good composer', you are not using the term 'good' in its moral meaning. For we can *aesthetically* praise something that we *morally* condemn. The movie *Triumph of the Will* , which was a Nazi propaganda film, is held to be a cinematic masterpiece. But it was obviously not something *morally* good. So it would be no contradiction to say '*Triumph of the will was aesthetically good, but morally bad.*' But, of course, if the aesthetic meaning of 'good' coincided with the moral meaning of that word, that sentence would be self-contradictory.

In this paper, when I use the terms 'good' and 'bad', I will be using them *only* in their moral sense. The claims that I make about these terms will usually not be true if these words are taken in their non-moral senses.

II. Leaving aside various niceties that will have no special bearing on the present inquiry, a string of verbiage qualifies as a declarative sentence if it is a conventionalized way of expressing a certain kind of belief. (In this paper, by 'sentence' I mean 'declarative' sentence.) Consider the verbiage 'snow is white'. If you utter this, you are producing a sentence, and this is ultimately because that verbiage is a conventionalized expression of a certain kind of belief – the belief that snow is white. [225]

So 'Stalin was short' is meaningful because it is a conventionalized way of expressing a certain kind of belief. *If* 'Stalin was evil' is meaningful, that is because it is a conventionalized way of expressing a certain belief. There can be no doubt that valuative sentences are conventionalized ways of expressing *attitudes.* 'Abortion is wrong' and 'Stalin was evil' are conventionalized ways of expressing some kind of negative attitude towards murder and Stalin.

Some case can be made that the attitudes expressed by sentences like 'abortion is wrong' and 'Stalin was evil' have *emotional* content. In fact, a

case can be made that those attitudes *are* emotions. (This is of course controversial; I will give a bare outline of a way to discharge this contentious point.) In effect, a case can be made for emotivism. Here is such a case. (For the record I believe that emotivism is correct. But I admit that the argument I'm about to present falls short of proving it, though it probably gives it some credence and will, in any case, set forth *a* reason why one might be an emotivist.)

An Argument for Emotivism

Suppose that you have no negative emotion towards abortion. In that case, you could not sincerely say 'abortion is wrong.' A precondition for sincerely saying 'x is good' or 'x is bad' is that one feel a certain way towards x – that one feel positively or negatively towards x.

One might think that one can sincerely say 'abortion is wrong' or 'feeding the hungry is good' and yet not have any emotion towards abortion or towards feeding the hungry. But this position is hard to sustain. Consider a being who was completely devoid of emotions. Ask yourself: could that being really see abortion as wrong? Obviously that being would be indifferent to the fact that abortions occur; for if it were not indifferent to that fact – if it were moved by that fact – then ipso facto it would have emotions. So it wouldn't care whether abortions were performed. If it didn't care, then how could it sincerely condemn abortion? How could it sincerely say that abortion was wrong? If a being was completely indifferent to whether abortions were performed – and obviously this is how an emotionless being would be --then it would be misrepresenting its state of mind in saying 'abortions are wrong' or 'I disapprove of abortions.' I simply don't see how the making of value judgments about x can be disentangled from the having

of emotions concerning x. (This is one of the controversial assumptions that I spoke of earlier.)

Admittedly, it might *seem* as though one can sincerely but emotionlessly say 'lying is wrong' or 'Hitler was evil.' But I think that this point of view is superficial. I think, in fact, that it confuses emotion with the occurrent manifestations of emotion. To be sure, one can sincerely say 'lying is wrong' or 'Hitler was evil' without having any *occurrent* feeling in connection with those words. But if one had *no* emotion in connection Hitler, then one would *ipso facto* be indifferent to Hitler and his deeds. And if one were truly *indifferent* to these deeds, could one really sincerely *condemn* them? Surely there is *some* intuitive pull to say 'no.'

Another *prima facie* objection the argument just given is this: "A psychopath – e.g. Ted Bundy – can *know* that what he is doing is wrong, but have no feeling towards his evil deeds, or having nothing but positive feelings towards those deeds. But if believing that e.g. rape is wrong just *is* having some kind of a negative emotion towards rape, then there would be no Ted Bundy's – there *would* be no people who *knew* that what they did was wrong, and yet had no negative feelings towards those deeds."

This objection has some force. But surely it is not conclusive. The emotivist can respond, and usually does respond, by saying that a psychopath is incapable of grasping the concepts of good and evil. Indeed, this is not only how emotivists characterize psychopaths: this is how many psychiatrists characterize them. According to this point of view, Ted Bundy and Jeffrey Dahmer knew that what they were doing was wrong only in a derivative sense: they knew that what they were doing was wrong in the sense that they knew that society disapproved of it.[226] They grasped society's disdain for their deeds, but they themselves didn't really grasp the wrongness of their deeds. (Whether this counter-reply is correct, I do not know. But it has some

credibility.)

Thus, there is at least *some* reason to hold that a precondition for sincerely uttering a valuative of sentence is that one have an emotion of some kind. (Admittedly, these reasons are inconclusive. *Vide* the two objections to it just considered.) If this is right, then a precondition for sincerely saying ‘x is good’ or ‘x is bad’ is that one have some kind of a positive or negative emotion towards x. It is plain, in fact, that in sincerely saying ‘x is good’ or ‘x is bad’, one is expressing one’s positive or negative emotions towards x. Further, it is clear that ‘x is good’ and ‘x is bad’ are conventionalized ways of expressing such emotions. There are no linguistic rules governing when to groan. So groaning, although obviously an expression of some kind of psychological state, is not a linguistic expression of such a state: for it is not regulated by linguistic rules. But there are obviously linguistic rules governing when to say ‘x is good’ and ‘x is bad’. There are linguistic conventions regulating the use of ‘good’ and ‘bad’. So in as much as these terms, in virtue of their functions within some conventionalized means of self-expression, embody positive and negative emotions, it follows that there are linguistic rules governing when to say ‘x is good’ and ‘x is bad’: it follows, in effect, that ‘x is good’ and ‘x is bad’ are conventionalized ways of expressing emotional attitudes, just as ‘x is tall’ and ‘x has a positive charge’ are conventionalized ways of expressing beliefs. (My statement that ‘x is good’ is a conventionalized way of expressing an emotion is obviously not free from controversy. I hope, not to have established it conclusively, but only given it enough credence to give some credibility to what follows.)

So (i) valuative sentences have emotional content and (ii) this emotional content is central to their significance: it is not incidental to the semantics of ‘x is good’ that, when sincerely uttered, it expresses a positive attitude towards x. On the contrary, the semantics of ‘x is good’ nearly enough

consists in that sentence's being a conventionalized way of expressing a positive attitude towards x. Obviously the same point applies mutatis mutandis to 'x is bad', 'x is evil', 'x is honorable' and so on.

Whatever the merits of the argument just given, it is clear that emotivism is to be taken seriously. Now for some reason, emotivism is held to be *incompatible* with realism. It is held that *if* emotivism is correct, then 'x is good' and 'x is evil' *do not* express facts about x – do not really *say* anything about x (they only *show* how the speaker *feels* about x). It is held that *if* 'x is good' and 'x is bad' express our emotions about x – our *feelings* about x – then they do not have any 'cognitive value', they do not *say anything objective about x.*

Evidently, it is thought that, in so far a sentence-token expresses how the speaker *feels*, then to that extent it provides no information about anything other than the speaker's emotions. We will presently see that this line of thought involves a serious non-sequitur -- it confuses what is *expressed* by a sentence token, on the one hand, and what is *said* by it, on the other.

III. The truth is that emotivism is opposed to realism *only if* a highly dubious thesis about emotions is correct; i.e. 'x is good' and 'x is bad' *do* make true or false statements about x *unless* a certain highly dubious thesis about emotions happens to be true. Put in positive terms, emotivism *is* a form of realism *if* a highly credible thesis about emotions – one held by Sartre, Robert Solomon, Merleau-Ponty, and various others – is correct.

That thesis is this: emotions are beliefs or judgments of a certain kind. I myself hold that this thesis is correct – in my view, emotions *are* beliefs of a certain kind: they are *valuative* beliefs. And I will outline an argument for

this thesis. (That outline will fall far short of proving that cognitivism is correct, but it will give that doctrine some credence.) Let us refer to this thesis as ‘cognitivism.’ So ‘cognitivism’, as I will be using this expression, is the thesis that emotions are beliefs of a certain kind.

The term ‘cognitivism’ has various meanings in philosophy; it sometimes refers to what I have been calling ‘moral realism’. In the philosophy of mind, it refers to various different, and not always compatible, doctrines. My use of this term is not meant to be consistent with any existing usage of it; it is entirely stipulative. To be maximally clear: by ‘cognitivism’ I mean the doctrine that emotions *are* beliefs of some kind.

At first, it might seem that cognitivism is an outrageous view:

“Admittedly emotions incorporate beliefs. But surely there is more to an emotion than just a belief. Surely there is more to my being angry at Mary than my just having a belief.” ***[227]***

But given some rudimentary and uncontroversial truths about emotions, it turns out to be much harder, I think, to make a case *against* cognitivism than to make a case *for* it.[228]

An Argument for Cognitivism

I will now offer *an* argument for cognitivism. Whether it proves cognitivism to be correct, I do not know. But it probably gives it enough credibility to make its consequences worth exploring.

First of all, an emotion is not *just* a feeling. If one has an object-less

feeling, one is not thereby having an *emotion.* Suppose one feels good but does not feel good *about* anything.[229] Such a state, surely, is not an emotion: it is a *mere* feeling. So emotions are *directed*. They have the property of 'intentionality': 'emotional consciousness is…consciousness *of* the world.'[230] In having an emotion, one is feeling a certain way *about* something. To love Mary is to feel a certain way *about* her; to be angry that Bob stole my bicycle is to feel a certain way *about* the fact that Bob stole my bicycle.

I should clarify what I mean by the term 'feeling'. When I say that, to have an emotion is to 'feel' a certain way about something, I don't mean that to have an emotion is to be in a state that has *phenomenal* content – in a state that has a *phenomenology*.

To say that a mental condition has 'phenomenal content' – or, equivalently, that it has a 'phenomenology' – is to say that there is *something it is like*[231] to be in that state. There is something it is like to smell coffee or taste mustard. So those conditions have phenomenal content. By contrast, there is not anything it is like to believe that two is the only even prime: so that mental condition – if 'condition' is the right word – has no phenomenal content: no phenomenology.

To be sure, some emotions – so-called *occurrent* ones – do have phenomenal content. If I see a rabid dog running towards me, I will feel a surge of fear; and that emotion has phenomenal content. But not all emotions are occurrent; and it is therefore not always the case that, in having an emotion, one's mental life has a certain phenomenology. For example, I am happy that my uncle's chemotherapy was a success. But it is not the case that, in virtue of having that emotion, I am at every waking moment in a state with a certain phenomenology. There are people who I currently love/hate who I

haven't even thought about in months, and in connection with whom I therefore haven't been in any phenomenally pregnant states for some time. So when I say that to have an emotion is to 'feel' a certain way about something, I am not using the term 'feeling' to denote an occurrent, phenomenally pregnant condition; I am using to denote an attitude of some kind. [232]

Once it is granted that emotions are essentially *directed* – more specifically, that they are *attitudes* of a certain kind -- it becomes very hard to see them as anything other than beliefs. Uncontroversially, my being angry that Bob stole my bicycle consists in my having a certain attitude – a negative one – towards a certain (putative) fact, viz. that Bob stole my bicycle. (Analogous remarks apply to all other emotions.) What is it to have an *attitude* towards a fact? One plausible answer is this: to have an attitude towards something is to see that fact thing in a certain light. And to see something in a certain light is, presumably, to see it as being a certain way -- as having certain properties. To have an attitude towards the Russian revolution to see it in a certain light; and this, in turn, is presumably to see it as being a certain way. To see the Russian revolution as being a certain way is to presumably to have some kind of view as to what properties the Russian revolution had, and is thus to have some set of *beliefs* about it. (Again, here I am making use o several debatable, but highly plausible, assumptions. First, I am assuming that to have an attitude towards something – to be angry at it or happy for it – is to see it 'in a certain light.' Next, I am assuming that to see something in a certain light is to see it as being a certain way. And I am also assuming that to see something as being a certain is to see it as having certain properties. One might contest these identifications. But there can be no doubt that there is some presumption in favor of each of them.)

This line of thought suggests that emotions are beliefs of *some* kind. But

it doesn't yet show that emotions are *valuative* beliefs. But once it is granted that emotions are beliefs of *some* kind, it is not hard to show that they must be *valuative* beliefs. (A valuative belief may *incorporate* a factive or non-valuative component. In order to believe that it was evil of Stalin to kill millions of people, I must have the non-valuative belief that Stalin killed millions of people. By a 'valuative' belief, I don't mean a belief that has *no* factive component; I mean a belief that involves the ascription of some positive or negative value to something. Valuative beliefs, it seems to me, typically involve factive or non-valuative components.)

Given any purely *non-valuative* belief, one does not, in principle, *have* to have any specific emotion in connection with it; and, in principle, one can have *no* emotion in connection with it. Consider the belief that Stalin killed thirty million people. We can imagine some sadist who had this belief, and *delighted* in it. (Maybe Stalin himself is an example of such a person.) And we can imagine someone else who had this belief and had *no* emotion, positive or negative, in connection with it – someone who was indifferent to Stalin's atrocities. (Maybe Molotov is an example of such a person.) Clearly, given any belief that has no valuative component, we can imagine a being that had no emotion, or had different emotions, in connection with that fact.

Given this, it is not possible that an emotion should be identical with a non-valuative belief. If my having emotion of type E were *identical* with having non-valuative belief of type B, then it would not be possible to have *different* emotions in connection with B and it would not be possible to have *no* emotion in connection with B. If my being *angry* that income is not evenly distributed were *identical* with my believing that income is not evenly distributed, then it would not be possible, even in principle, to have *no* emotion in connection with the belief that income is unevenly distributed, and it would not be possible, even in principle, to have some emotion *in*

place of anger in connection with that belief.

Also, given that one can have different emotions, or *no* emotion at all, in connection with a given non-valuative belief, it follows straightaway that emotions do not *supervene* on non-valuative beliefs. Given that one can have *different* emotions, or even *no* emotion, in connection with any type of non-valuative belief, it follows, by the definition of 'supervenience', that emotions do not supervene on non-valuative beliefs.[233]

So, if the having of emotions consists in the having of beliefs, it cannot be identical with, or supervene on, the having of non-valuative or purely factive beliefs. It follows that if emotions are beliefs, they must be valuative ones.

This position is borne out by consideration of individual emotions. Suppose that Bob stole my bicycle and that I am angry that Bob did that. That emotion obviously consists in my having some kind of an *attitude* – a negative one – towards the fact that Bob stole my bicycle. And having such an attitude would seem to consist in my seeing that fact in a certain light: in my seeing it as a *bad* thing, as having negative value. Suppose that Mary just won the Nobel prize and that I am happy about that. My being happy that Mary won the Nobel prize consists in my having a certain attitude towards that fact. And my having that attitude would seem to consist in my seeing that fact in a certain light: in my seeing it as a *good* thing. My fearing that Mary will kill me consists in my believing it to be at least not entirely out of the question that Mary might kill me, and in my feeling a certain way about – having a certain attitude towards – that fact: in my seeing her killing me as something that, if it were to happen, would have some kind of negative value.

There is a slightly different way to give support to this line of thought. For me to be *angry* that Bob stole my bicycle, I must obviously *believe* that Bob stole my bicycle. But, in principle, I could *believe* that Bob stole my bicycle

without having any emotion in connection with that fact. (Maybe I want Bob to have a bicycle so badly that I am *glad* he stole my bicycle. Maybe, as part of some elaborate plot to frame Bob, I *wanted* Bob to steal that bicycle. Would Mr. Spock be angry if he knew that Bob stole *his* bicycle? What about the Dalai Lama?) So my being *angry* that Bob stole my bicycle cannot consist *merely* in my believing that Bob stole my bicycle. But if what we said about emotions is correct, they are beliefs of some kind. So what kind of a belief could the aforementioned anger be? Well, there is no purely *factive* belief – no purely non-valuative – belief that this anger could consist in. For given any purely factive belief, I don't *have* to have any emotion in connection with it and I could, in principle, have very different emotions in connection with it. So *if* – what we've seen some reason to hold – emotions, are belief, they must be valuative beliefs. So my being angry that Bob stole my bicycle cannot consist in my having this or that non-valuative belief. So it must consist in my having some valuative belief. And when we examine this anger, it seems to consist in my seeing Bob's theft of my bicycle as having a certain negative *value* – as being a *bad* thing.

Let us sum up. Emotions are attitudes towards things. To have an attitude towards a thing is to see it in a certain light. To see something in a certain light is to see it as having certain properties. To see something as having certain properties is to have a belief about it. Thus emotions are beliefs. (But, of course, not all beliefs are emotions. My belief that 2+2=4 is not an emotion.) Emotions cannot be identical with, or supervene on, non-valuative beliefs; for any non-valuative belief can be associated with different emotions or with no emotion. Therefore they must be valuative beliefs. So here we have a plausible, purely theoretical argument for the position that emotions are valuative beliefs. Consideration of individual emotions provides a kind of empirical support for this purely theoretical argument. My being happy that

Mary did such and such seems to consist in my seeing her doing such and such in a positive light, and thus in my believing it to have some kind of positive value.

The temptation is very strong to say that, although emotions *incorporate* beliefs, they are not *just* beliefs: they can be decomposed into doxastic and non-doxastic components. So, according to this view, my negative attitude towards Bob's theft of my bicycle – i.e. my anger in connection with this fact – is decomposable into a belief (viz. that Bob stole my bicycle) *plus* something other than a belief. I cannot state with certainty that this view is wrong, but there is no *obvious* way to effect this decomposition (I challenge anyone to try) and, beyond that, there is some reason to think that it cannot be done.

Suppose we say that this anger consists in (i) my believing that Bob stole my bicycle *plus* (ii) some non-directed (object-less) bad feeling. Well, in that case, I plainly don't have a bad feeling *about* Bob's stealing my bicycle; so I do not have a negative attitude *towards* this fact – and I don't have any kind of *emotion* towards it.

On the other hand, suppose we try to decompose my anger into (i) my believing that Bob stole my bicycle *plus* (ii) some directed (representational) bad feeling. In that case, we have to ask: what is it to have a *representational* bad feeling? Presumably, it is a feeling which *represents* something *as* being a certain way: it is my having a bad feeling *to a certain effect*. But such a thing is nothing other than a belief.

To sum up, the 'de-cognitivizing' of emotion – the attempt to recover the quite understandable intuition that there is *more* to an emotion than just a belief – is not such an easy enterprise: there is no obvious way to carry it out. And this, in itself, constitutes a kind of argument, if only a weak one, for cognitivism.

IV. Obviously, we have not proven that cognitivism is true; our argument was replete with controversial assumptions. But we have, I think, shown that it is a doctrine to be taken seriously. Now I want to show that *if* cognitivism is true, *then* emotivism is a form of moral realism. Of course, this 'if-then' statement is of interest precisely because the 'if' part may be true -- cognitivism may be accurate; so emotivism may well be a form of realism.

According to emotivism, when one says 'x is good' or 'x is bad' one is expressing an emotion towards x. Now it is plain that *if* cognitivism is true, then when one says 'x is good' or 'x is evil' one is expressing a belief. (Of course, when one says 'x is tall' or 'x has a positive charge' one is expressing a belief.) So *if* cognitivism is correct, then emotivism is a form of realism, and is thus --- contrary to popular opinion – in no way opposed to it.

Obviously people who see a conflict between emotivism and realism think that in so far as 'x is good' expresses how the speaker *feels* about x, to that extent it doesn't say anything about x.[234] The truth is that, even if 'x is good' expresses how the relevant speaker *feels* about x, that does not by itself mean that 'x is good' to any extent *fails* to make a true or false statement about x. To believe otherwise is to fail to distinguish to between what a sentence *expresses* and what it *says.*

The sentence 'Smith is short' is about Smith; it is obviously not about the person who produces that sentence (unless, by sheer accident, that person happens to Smith). Suppose Bob says 'Smith is short'. Bob is obviously not saying anything about himself. But he *is* obviously *expressing* a belief that he has. So, in effect, Bob is *showing* something about himself – he is *showing* that he *believes* Smith to be tall – even though he (Bob) is not *saying* anything about himself. One cannot say anything without expressing some belief that one has – without *showing* something about one's own beliefs and,

therefore, about oneself. Now suppose – what we've seen some reason to be the case – that one is expressing an emotion in saying 'x is good' or 'x is bad'. In other words, suppose that emotivism is correct. Given only that emotivism is correct, it doesn't follow that, in saying 'x is good' or 'x is bad', one is not also making a true or false statement about x. For if cognitivism is correct, then in having an emotion, one is also having a belief; and in that case, if sentences like 'x is good' and 'x is bad' serve to express emotions, that means they serve to express *beliefs.* And if that is the case, then 'x is good' doesn't operate differently from 'x is tall': both sentences express beliefs and both make objective statements.

References

The Cambridge Dictionary of Philosophy general editor Robert Audi Cambridge 1995

Ayer, Alfred Jules

[1] *Language, Truth, and Logic* Dover 1952

Blackburn, Simon

[1] *Spreading the Word* Clarendon Oxford 1984
[2] "Supervenience Revisited" in *Essays in Quasi-Realism* Oxford 1993

Cheshire Calhoun and Robert C. Solomon editors
What is an Emotion? Oxford University Press 1984

Hume, David
A *Treatise Concerning Human Nature* Edited, with an Analytical Index, by L.A. Selby-Bigge; Second Edition, with text revised and variant readings by P.H. Nidditch. Oxford at the Clarendon Press 1978

Merleau-Ponty, Maurice
The Phenomenology of Perception Routledge 1962

Sartre, Jean-Paul
The Emotions: A Sketch of a Theory 1939
The Psychology of the Imagination translated by Bernard Frechtman Washington Square Press 1948

Stevenson, C.L.

[1] "The Emotive Meaning of Ethical Terms" in *Logical Positivism* ed. A.J. Ayer The Free Press 1959

NO SUCH THINGS AS MENTALESE

Abstract

According to Jerry Fodor, for someone to grasp a proposition – to have a thought – is just for some sentence of an internal language to be tokened. This internal language is thus the 'language of thought'. In this essay I will provide some reasons to think that this idea is untenable. For someone to know a language L, that person must know the rules giving meaning to the expressions of L. (Admittedly, this idea may be questioned; but reasons are provided for it.) So for someone to know the language of thought, he must know the rules which assign meaning to the expressions belonging to that language. Now we know from Tarski that no language can express the rules that give its own expressions meaning. So if one is to grasp the semantical rules for some language L, one must either grasp them (i) in some 'Meta'-language (that comprises L as a proper part of itself) or (ii) one must grasp them in some altogether non-linguistic form. Someone who holds that all thinking consists in the tokening of symbols cannot take path (ii). So he must take path (i). This means that if one is to understand a language L, one must understand the relevant meta-language and, therefore, must grasp the rules that give meaning to the expressions of this meta-language. But by Tarski's theorem, these rules cannot be given in the meta-language itself. So if one is to grasp these rules – and, therefore, if one is to *understand* the meta-language – these rules must be given to one either (i') in some meta-meta-language or (ii') in some non-linguistic form. If we adopt path (ii') we are abandoning the idea that thinking consists in the tokening of symbols, and thus abandoning the idea of a language of thought. On the other hand if we take (i') we are launched on a vicious regress. Ways of blocking this regress, without abandoning the idea that there is a language of thought, are considered and, ultimately, deemed inadequate.

I. Fodor's Thesis and My Counter-Thesis

According to Jerry Fodor, there is a 'language of thought'. Fodor does *not* mean by this that English speakers think *in* English, that Albanian speakers think *in* Albanian, and so on. (In fact, Fodor does *not* hold this. He thinks that *in order* to know a natural language – e.g. English, Albanian – one must *already* be able to have thoughts; and that, consequently, thinking does not ultimately take place in any natural language.) What Fodor means is that there is some *innately known* language – some language that doesn't have to be learned – such that all thinking consists in the tokening of expressions of this language.

Fodor's argument is, very roughly, this. (i) Thought consists in, or at least involves, the entertaining of propositions. (ii) All representation involves some kind of *medium*. (iii) Any medium capable of representing propositions is *ipso facto* a language. (iv) Therefore the medium in which thought takes place is a language of some kind. (v) Therefore there is a 'language of thought': thought consists in the tokening of expressions of some language. (Fodor refers to this putative language as 'Mentalese'.) [235]

In this essay, I will provide just *one* reason to hold that there couldn't possibly be a language of thought; that thinking could not, in any literal sense, consist in the tokening of expressions.

I should make it clear that I am *not* approaching this question from an anti-cognitive science perspective. In fact, I am biased *in favor* of a cognitive-scientific conception of mental activity. I have no doubt that the basic postulates of cognitive science are correct: virtually all of our mental productions involve the existence of sub-conscious and, what is more, sub-*personal* mentation, and this mentation consists in the occurrence of *inferences*

or *computations*. So, for example, mediating between the impact of various of photons one's retinae, on the one hand, and the subsequent having of visual experiences, on the other, are not only various physiological processes, but also various *cognitive* processes: data is integrated; inferences are made; conclusions are drawn – all at a level far beneath that of personal awareness and, indeed, necessarily inaccessible to introspection. So, just to be clear, my purpose is *not* to undermine cognitive science; it is rather to eliminate an erroneous way of thinking about cognitive scientific inquiry.

II. A Problem for Fodor's Thesis: Its Need for Our Knowledge of an Infinite Hierarchy of Meta-Languages

Any English-speaker is capable of understanding sentences of English that he hasn't encountered before. (You can understand the sentence 'feathered pigs play tennis with zeal' even though you've probably never encountered it before.) It is not an *accidental* fact that any English-speaker can understand sentences he has never encountered before: having this ability seems to be *constitutive* or *definitive* of being an English speaker. Suppose someone could only understand some finite stock of sentences – e.g. 'John is tall', 'Sam is fat', and 'cats like dogs'. (Actually, later on I will try to show that, unless one could understand an infinite number of English sentences, one could not, ultimately, understand *any*. But for now let us operate on the fiction that one *could* understand only some finite number of English sentences.) Such a person – though he might understand *some* English sentences – could nevertheless not really be said to know English. For knowing English involves being able to understand – to assign meaning to – sentences one has not encountered before: if one really knows English, then the set of English sentences one can understand is, in some sense, 'unbounded'.

So knowing a natural language involves being able to understand sentences

that one has not previously encountered, and whose meanings one has therefore not *memorized*. Given this, it seems reasonable to suppose that, in order to know a natural language, one must know some set of *rules* assigning meaning to expressions of that language.

These rules are, presumably, of two types: rules that assign meaning to *simple* expressions (e.g. 'cat', 'dog', 'red', 'Sam'); and those that assign meanings to complex expressions on the basis of the meanings of the simple expressions composing them ('recursive' rules). You can understand 'Bill Clinton smokes cigars' because (a) you know what 'Bill Clinton', 'smokes', and 'cigars' means and (b) you know how these meanings combine, as it were, to form the meaning of the sentence in question.

Now if there is a language of thought, one could not know this language simply by *memorizing* the meanings of its expressions. There is no pre-set limit to the thoughts a cognitively normal human being can have. It is not as though there is some finite stock of thoughts – e.g. *snow is white, grass is green, girls play tennis,* and *baboons write poetry.*- such that one can have *only* those thoughts and no others. The set of thoughts one can have is, at least along certain dimensions (though not all, to be sure) 'unbounded'. If one can grasp certain thoughts, then one can generate new ones by combining and iterating those thoughts. So one cannot understand the language of thought, supposing that there is such a thing, by *memorizing* the sentences belonging to that language; for there are infinitely many of them; and, presumably, an infinite list of sentences cannot be memorized.

Also, thought has a 'non-atomistic' character. If one can have one thought, one can have many others; any given thought has 'constitutive' connections to other thoughts. (It seems impossible that a creature should be able to have *just* the thought that the interior angles of a triangle add up to 180. For if one can have one thought, one can have many others: many other thoughts about

arithmetic, geometry, rigidity, direction, causality, and so on. For example, it is impossible that you should be able to grasp the proposition that x, y, and z add up to 180 without grasping the proposition that x, y, and z add up to 179 or 181; for the thought that x, y, and z add up to 180 is an essentially *contrastive* thought: it can exist only in a space of other arithmetical beliefs. And there doesn't appear to be any definite limit to the number of beliefs required to generate the appropriate 'space'.) So it seems a bit unlikely that a being capable of understanding the language of thought – capable of understanding 'Mentalese' – can simply have memorized the meanings of Mentalese sentences; for any one thought is, apparently, embedded in a matrix of other thoughts: a matrix that would appear to extend indefinitely in at least some directions. Therefore, it seems reasonable to presume, anything capable of understanding Mentalese must know some set of *rules* assigning meaning to these sentences.

So we must ask: given a being that understands Mentalese sentences, how are the rules assigning meaning to these sentences *given* to that being? How are they represented to it? Trivially, either these rules are given to that being through a language or *not* through a language. If they are *not* given to it through a language, then *ipso facto* that being can grasp propositions that are *not* embodied in an expression of a language, i.e. thought does *not* (at least not in general) consist in the tokening of expressions of some language. So anyone who believes in Mentalese – who believes that thought consists in the tokening of expressions in a language – can*not* hold that the rules assigning meaning to Mentalese sentences are given in some non-linguistic form. So, it follows trivially that these rules must be given in some linguistic form: in some language.

Trivially, either these rules are either given in Mentalese or they are not. Now Tarski proved that *no language can contain the rules assigning meaning*

to the expressions of that language.[236] As it is sometimes put, no language can contain its own 'truth-predicate'. Put yet another way, given any language L, the rules giving meaning to sentences of L cannot *themselves* be expressed in L. This is a theorem of mathematics, not a psychological or linguistic hypothesis: as a matter of mathematics, it is not possible for a language to be capable of expressing the rules that assign meaning to its own sentences.

So anything capable of understanding Mentalese must either (i) grasp the rules giving meaning to Mentalese expressions through some language *other* than Mentalese or (ii) it must grasp these rules in some *non-linguistic* form. We've already seen that anyone who believes in a 'language of thought' cannot accept (ii). So anyone holding that there is a language of thought –that thoughts are tokenings of expressions - must accept (i); he must hold that one knows the rules giving meaning to Mentalese sentences by virtue of grasping sentences of some language L' (distinct from Mentalese). So anything capable of understanding Mentalese must grasp some language L' *distinct* from Mentalese (though L' may contain Mentalese as a proper part of itself). Now L' must be rich enough to give meaning to all the expressions of Mentalese; so it must have a similarly 'unbounded' nature. So one cannot understand L' by virtue of *memorizing* the sentences belonging to it. One must know the *rules* assigning meaning to these sentences. By Tarski's theorem, these rules cannot be stated in L' itself: they must be given in some *other* language L''. But, by an argument exactly analogous to that just stated, one cannot understand L'' by memorizing the sentences belonging to it; one must know the rules assigning meaning to these sentences. These rules, by Tarski's theorem, cannot be stated in L'', but must be stated in some other language L''' – and so on *ad infinitum*. So with regard to the idea that rules can be grasped only in some linguistic form, this idea leads to a vicious regress: it involves the idea that, to grasp a single language, one must grasp an infinite hierarchy of languages. But it is

unreasonable to suppose that everyone knows an *infinite* hierarchy of languages; indeed the idea is probably absurd. So if one is to know any language – at least any language rich enough to mediate the potentially unbounded set of thoughts any human can have - then propositions are not always given in linguistic form: the grasping of propositions does *not* consist, at least not universally, in the tokening of expressions.

So the idea that there is a language of thought is self-defeating. If there is a language of thought, then to understand the sentences of this language, certain rules, certain propositions (viz. those assigning meaning to said sentences), must be given to one in *non*-linguistic form; so that the grasping of does *not* consist in the tokening of expressions of some language. But if the grasping of propositions – the having of thoughts – does not consist in the tokening of expressions of some language, then the hypothesis that there is a language of thought is false: for that hypothesis just *i s* the idea that the grasping of propositions consists in the tokening of 'Mentalese' expressions.

III. Fodor's Response and My Counter-response

Fodor's Response

At this point, I must address several objections that might be leveled against my argument. One of these is due to Fodor himself, who was responding to an earlier draft of the present article. I quote:

[Y]ou don't have to represent the rules of a language that you are able to use; all that's required is that you are so constructed as to conform to the rules. (Cf the way computers are `built to use' their machine language.) Of course, it's plausible that you DO internally represent the rules of (say) English; but that's not because English is a language; it's because it's a language that you LEARN; and it's quite plausible that to learn a

language is to come to internally represent its syntax and semantics. (Presumably in `Mentalese').

Here, in essence, is what Fodor is saying. You mentally represent the rules of English (in Mentalese); but you don't represent the rules of Mentalese themselves. Hence the regress I discussed is blocked. How, then, do you (or the relevant part of your mind) understand Mentalese expressions? Well, you are simply 'built' to understand them; you are structured in such a way that, when confronted with a token of Mentalese, you arrive at an understanding of its meaning *without* applying knowledge of the rules which give that expression meaning. So if you are confronted with the Mentalese sentence **if snow is white, then something is white** you *don't* apply some knowledge of rules to assign meaning to the 'snow', 'white', etc. and then use some knowledge of recursions to get the meaning of the whole sentence: you *don't* do that, because that would involve your having a knowledge of the rules giving meaning to Mentalese expressions – which rules cannot be given in Mentalese itself, given Tarski's result – and the regress I discussed gets started. Rather you ascertain the meaning of **if snow is white, then something is white** by virtue simply of being 'so constructed' that you automatically – i.e. in some way not involving knowledge of linguistic rules, not involving computation: indeed, not involving *cognition* (for surely cognition is mentation that involves consideration of rules) – arrive at a knowledge of its meaning.

My Counter-Response

This position is untenable on a number of levels. But, ultimately, I think, there are two problems with it. *(1):* Given a linguistic expression E, if you don't have to *figure out* what E's meaning is on the basis of any constituent expressions that it might have, then *ipso facto* E is a *primitive* expression, at

least from your point of view. Now if Fodor's counter-objection (quoted above) is correct, then you never *figure out* what Mentalese expressions mean. (You don't have to *compute* their meanings.) *A fortiori* given any Mentalese expression, you don't figure out what they mean on the basis of the meanings of any expressions that might compose them. So, from the point of view of the person (or sub-personal agency) that has to understand mentalese expressions, each of these expressions is *semantically primitive* (if they are semantic at all). So the Mentalese sentence *if snow is white, the something is white* is primitive; it is *not* a compound of the Mentalese sentences *snow is white* and *something is white*. Now Fodor holds that to grasp a proposition just *is* to grasp a Mentalese sentence. So he must also hold that to grasp relations among propositions – e.g. the relation that holds between *snow is white* and *if snow is white then something is white* – just is to grasp relations (of meaning, presumably: *semantic relations*) among sentences. But if each Mentalese sentence is semantically *primitive*, then there are no semantic relations among them; which implies that any creature thinking in Mentalese could (indeed must) grasp *if snow is white then something is white* without grasping how the constituent sentences in this complex sentence relate to the latter. But this is absurd.

(2) With regard to the idea that one could *understand* an expression without knowing the rules that assign it meaning: that idea, I submit, is incoherent. Suppose that S is some sentence and that you know what proposition it encodes, but that you arrived at this knowledge *without* applying some set of semantic rules: in that case S is not, at least not from your point of view, a bit of *language*. In any case, this is what I will now argue.

I admit that both of these claims are far from self-evident. So I would now like to elucidate and defend them. Let's start with (1).

(1) Fodor's response incompatible with our ability to grasp logical relations holding among propositions

First I'd like to illustrate the idea of a 'semantic primitive'. Suppose that the following is a long sentence of Finnish:

(i) xyz

And suppose that this sentence means

(ii) If pigs are the only flying creatures, and only flying creatures have horns, then only pigs have horns.

You now know what proposition (i) encodes. But you don't really *understand* the sentence, at least not in the sense in which a Finn does. For you have no idea what contribution the various components of (i) make to the proposition expressed by that sentence. So at least one of the dimensions of (i)'s meaning that is known to any Finn is unknown to you. We can characterize this difference by saying that *for you* (i) is a primitive expression, whereas for a Finn (i) is a highly complex expression. So even though the *proposition* expressed by (i) is complex, nevertheless the *sentence* itself is, from *your* point view (though not a Finn's), semantically primitive: for, from your point of view (though not a Finn's) its meaning is not to be arrived by performing a computation on the meanings of simpler expressions.

Now suppose that Fodor is right: suppose that you (or the relevant sub-personal agency within you) doesn't understand Mentalese sentences by consulting any set of *rules*; that you are simply 'so constructed' that you understand them. (Incidentally, the understanding of Mentalese sentences, if there are such things, occurs at an entirely sub-personal level. So, strictly speaking, it is incorrect to speak of *you* processing such sentences; for this term

obviously refers to a *person*, and not to a network of sub-personal computational processes. But for brevity's sake, I will speak of 'you' grasping, processing etc. Mentalese sentences. Simply realize that, in these contexts, I am using the term 'you' to refer to some sub-personal entity *within* you.) This means that, given any Mentalese sentence S, you don't *compute* its meaning. In that case, no Mentalese sentence that you come across is semantically *complex*, no matter how complex the proposition that it encodes.

Given this, consider the *English* sentence

(i) If snow is white, and all cows eat snow, and all things that eat white things like grass, then cows like grass.

It is pretty obvious that you understand (i) by virtue (inter alia) of understanding the sentences

(ii) snow is white

(iii) all cows eat snow;

(iv) all things that eat white things like grass;

and

(v) cows like grass.

It is pretty obvious, then, that you arrive at the meaning of (i) by, in some way or other, 'integrating' the meanings of these other sentences. And, of course, your 'integrating' these meanings consists in your applying some set of

rules that are defined over these meanings. In light of this, consider the *Mentalese* sentence:

(i') If snow is white, and all cows eat snow, and all things that eat white things like grass, then cows like grass.

One would presume that you can understand (i') by virtue of understanding the Mentalese sentences

(ii') snow is white

(iii') all cows eat snow;

(iv') all things that eat white things like grass;

and

(v') cows like grass.

But if Fodor is right, this is not the case. If Fodor is right, (i') is not – at least nor from your point of view: from the point of view the person who has to *figure out* what (i') means - *semantically* complex; its meaning is not to be *computed* in any way. For computation involves consideration of rules, and Fodor's way of blocking the regress discussed above is to say that one understands Mentalese sentences *without* consulting any rules. A fortiori, as far as you can tell –- for, if Fodor's objection is correct, you have no access to the recursions relating (i') to (ii')-(v') -- (i')'s meaning is not a function of the meanings of (ii')-(v').

But this idea is multiply incoherent. For Fodor – for anyone who believes that there is a language of thought - grasping the proposition *snow is white* simply *is* grasping the Mentalese sentence (ii') **snow is white**, and grasping the proposition *If snow is white, and all cows eat snow, and all things that eat white things like grass, then cows like grass* simply *is* grasping the Mentalese sentence (i') **If snow is white, and all cows eat snow, and all things that eat white things like grass, then cows like grass.** For Fodor, you have no way of grasping the propositions other than by grasping these sentences of Mentalese. (This is what the whole 'language-of-thought' hypothesis amounts to.) Now there's obviously no way to grasp the proposition (i) without grasping (ii)-(v): your grasp of (i) is surely 'built up' in some way from your grasp of (ii)-(v). For Fodor, grasping (i) simply *is* to grasp (i'). So, if Fodor is right, one cannot grasp (i') without grasping (ii')-(v'); and one's grasp of (i') is 'built up' in some way from one's grasp of (ii')-(v').

If grasping propositions simply *is* grasping sentences of Mentalese, then grasping relations among propositions consists in grasping relations among Mentalese sentences. What kinds of relations among Mentalese sentences? Presumably, relations of meaning: *semantic* relations. But if Fodor's objection is correct, you have no way of grasping such relations. Each Mentalese sentence is, for you, a semantic primitive. Your grasp of the Mentalese sentence **(i')** does not involve your grasping the Mentalese sentences (ii')-(v'). Your grasp of the *proposition* (i) is identical with your grasp of **(i).** So, if Fodor is right, your grasp of the proposition (i) does not involve your grasping the propositions (ii)-(v): you grasp the proposition *If snow is white, and all cows eat snow, and all things that eat white things like grass, then cows like grass* without grasping the propositions *snow is white*; *all cows eat snow*; and so on. But that is patently absurd.

Let us sum up what we've said so far. If the language of thought hypothesis

is correct, then to grasp a proposition is simply to grasp the Mentalese sentence representing that proposition. So to grasp relations among propositions is just to grasp relations among sentences: relations of meaning, presumably. Now Fodor's way of blocking the regress I mentioned earlier is to say that one doesn't ascertain the meanings of Mentalese sentences by performing computations, i.e. one does *not* derive the meanings of these sentences from the meanings of their parts. One is simply 'so structured' that one understands these sentences. So one does not understand such a sentence by taking in the meanings of its parts and, on the basis, computing the meaning of the whole sentence. Consequently, each of these sentences is, from the point of view of the person apprehending them, a semantic primitive. So no two Mentalese sentences have any semantic connection to each other, at least not as far as the person (or sub-personal agency) comprehending these sentences can tell. Now if grasping propositions is identical with grasping sentences of Mentalese, then grasping relations among propositions is identical with grasping semantic relations among Mentalese sentences. It follows that, if grasping propositions were identical with grasping Mentalese sentences, then one could grasp e.g. *if snow is white, then something is white* without grasping the proposition *snow is white*. But this is absurd.

A Fodorian might counter-respond as follows:

It is obviously true that if you grasp the proposition *if snow is white, and all cows eat snow, and all things that eat white things like grass, then cows like grass*, then you must grasp the propositions *snow is white*, *cows like grass*, etc. And, as you've pointed out, I hold that to grasp propositions simply *is* to grasp (tokens of) Mentalese sentences. So it does, admittedly, follow from my position that for someone to grasp relations of meaning among propositions is for him to grasp relations of meaning among sentences. But *these relations*

need not be recursive relations. For me to grasp the Mentalese sentence ***if snow is white, then something is white*** I must obviously grasp the Mentalese sentences ***snow is white*** and ***something is white*** and I must also grasp the relation that these two sentences bear to the sentence ***if snow is white, then something is white***. But this relationship is not a *semantic* relationship; it is just a logical relationship. In other words, my grasping this relationship doesn't consist in my knowing the recursion by which ***if snow is white, then something is white*** is assigned meaning on the basis of the meanings of its constituents; my grasping this relationship consists wholly in my grasping some relationship among propositions: some non-linguistic relationship.

This is precisely the view that a Fodorian cannot take. For the Fodorian, the grasping of logical relations coincides with the grasping of semantic – indeed, of recursive –relations. To grasp ***if snow is white, then something is white*** one must obviously grasp, first, the two propositions contained in it and, second, the way in which these propositions combine to form the larger proposition. For Fodor, to grasp a proposition is to grasp a sentence; so to grasp a complex sentence (one composed of other sentences) one must grasp, first, its constituent sentences and, second, the way in which these combine to form the larger sentence. Now to grasp the way in which the meaning of a complex sentence is determined by the meanings of constituent sentences – to grasp this *is* to grasp a recursive relationship among expressions. So there is no way for Fodor to avoid saying that one grasps recursions among sentences of Mentalese. Given Tarski's result, these recursions cannot be given in Mentalese itself; so they must be given in some meta-language. And the vicious regress we spoke of earlier starts again.

(2) Fodor's response incompatible with what it is to know a symbol's meaning

There is another problem with Fodor's response (this is connected with – is perhaps a different facet of – the problem just discussed). I begin with a simple question: can one understand a linguistic expression without knowing the rule that gives it its meaning? At first it seems not. If I hear the sentence 'snow is white', I understand that sentence just in case I know some biconditional like: 'snow is white' means *snow is white*. Fodor (in the quotation above) is saying that, where Mentalese sentences are concerned, I do not know such biconditionals. So I know the meaning of the Mentalese sentence **snow is white** without knowing any biconditional like **snow is white** means *snow is white*. So I *don't know* that **snow is white** means *snow is white*. So I don't know what **snow is white** means…? In that case, it would seem, a Mentalese token of **snow is white** would fall on deaf ears, as would (by an exactly analogous argument) every other token of a Mentalese sentence.

In response, Fodor would presumably say that one doesn't have to *interpret* Mentalese sentences: they wear their meanings on their sleeves; their meanings are 'self-presenting'; perhaps they simply *are* meanings and therefore don't have to be interpreted. (*Meanings* are not interpreted; things that *bear* meanings are interpreted: to interpret something is to figure out what meaning it bears. So meanings are not the sort of things that are interpreted.) To put it less metaphorically, the idea would seem to be this:

You don't *interpret* your visual perceptions. You just *have* them. They wear their meanings on their sleeves. Similarly, you don't *interpret* Mentalese sentence-tokens; for they were *their* meanings on their sleeves.

It may be true that we don't interpret visual perceptions. They seem to have a certain transparency. They are not (let us suppose) ciphers which have to be

decoded. But (supposing that this is the case) why is this? Because, presumably, the perception is *itself* a presentation of some kind of content. In having a perception, it is the *content* itself that is given to one, rather than some cipher which *represents* that content. In having a perception whose representational content is R, one is given, not some cipher which represents R, but R itself. (Of course, people differ as to what R is. Some people say it is a sense-datum. Others say that it is a completely mind-independent object. But, it seems, there is general agreement that the sense-perception is not a giving of something which *represents* this content – whatever it may be – but a giving of the content itself.) The idea suggested by the objector is that Mentalese sentences have the same kind of transparency as perceptions: when a Mentalese sentence is tokened in one's mind, the sentence *itself* is not given: only the thing that the sentence *represents*. But in that case the sentence drops out of consideration altogether. If there is nothing that has to be i*nterpreted*, then there is a *direct* grasp of meaning. But if there is a *direct* grasp of meaning, then the grasp of meaning in question is not mediated through any kind of *symbol*; and if it is not mediated through any kind of symbol, then *a fortiori* it doesn't involve any kind of *linguistic* symbol. So the idea that Mentalese expressions 'wear their meanings on their sleeves' – that they don't have to be interpreted – is self-defeating. A 'self-interpreting' symbol is not a symbol: if no interpretation is required to access meaning, then *ipso facto* no symbolic, and so no linguistic, intermediary is involved.

Where nothing has to be interpreted, there is no intermediary. Where there is no intermediary, there is no symbol and, a fortiori, no sentence and, therefore, no language.

This same point may be accessed from a slightly different angle. We don't want to say that *everything* representational is linguistic. We don't want to say that perceptions, mental images, attitudes etc. are *linguistic*; for that would

trivialize the idea of something being linguistic: of its being representational in a *semantic* way. It seems that at least part of the reason why my perception of the cat is not identical with some *symbol's* being tokened in my mind is that the perception doesn't have to be interpreted: what is given to me is not something which *bears* content; what is given to me is the content itself: what is given to me is not something which represents a cat, but must be deciphered; what is given to me is the cat itself. Now if I see the words 'the cat is on the mat', what is initially given to me is something which *bears* content, but which is distinct from that content. (The words 'the cat is on the mat' are presumably distinct from the proposition *the cat is on the mat* – unless we accept some kind of an extreme nominalism -- and the latter is the semantic content of the former.) In order to get at that content itself, I must, as it were, penetrate the thing which bears the content to get at the content itself; in effect, I must perform some act of interpretation. Now if nothing has to be interpreted – if the content itself is given, rather than something which bears that content – then we are not dealing with *symbols*. If *things* themselves are given to the subject, not representations thereof, then we are not dealing with symbols of things; and if we are not dealing with symbols of things, then we are not dealing with language. As we've seen, Fodor must deny that Mentalese sentences are *interpreted* (otherwise his view involves a vicious regress). And we've seen that, if Mentalese sentences are not interpreted, but are nevertheless understood, that is because the tokening of the Mentalese sentence is *identical* with the grasping of some meaning. So the tokening of such a sentence constitutes a *direct* grasp of meaning: the meaning itself is given, not some intermediary. But when we are dealing with a situation where there is no intermediary between the subject and the propositional content, then we are ipso facto not dealing with a *linguistic* situation. To sum up: if Mentalese sentences don't have to be *interpreted* that is because each tokening of a Mentalese sentence is an unmediated grasp of

meaning. But if a grasp of meaning is not mediated, then ipso facto there is no *symbolic* medium – no symbolic intermediary - between subject and the meaning grasped and, consequently, no sentence at all. So Fodor's position is self-defeating.

Saussure pointed out that, where natural language is concerned, the relation between sign and meaning is arbitrary. ('Snow' *could* mean *grass* or *omelet.*) Let us examine Fodor's hypothesis from the standpoint of Saussure's insight. Trivially, the relation between a linguistic expression is either arbitrary or it is non-arbitrary. So – what is also trivial - if there is such a language as Mentalese, the relations between its sentences and their meanings is either arbitrary or non-arbitrary. Let's consider each case. *First Case:* suppose the relation is non-arbitrary. In other words, suppose that any Mentalese symbol *has* to mean what it does. So **snow is white** *has* to mean *snow is white*; etc. Now what would make the relation between a symbol and its meaning non-arbitrary? Again, a comparison with the case of perception is helpful. The relationship between my current visual perception and its content is not arbitrary; is not a matter of convention or artifice. It is *constitutive* of this perception that it is a perception of a black computer on a brown desk, etc. (By contrast, it is not constitutive of the following ink-mark – 'grass is green' – that it means *grass is green.*) The reason is that the perception is not the presentation to me of something which *mediates* between representational content and myself; rather it is the presentation of representational content itself. So the non-arbitrary connection between the perception and its representational content is due, specifically, to the fact that the perception is a presentation, not of a symbolic medium, but of propositional content itself. So if the relationship between **snow is white** and *snow is white* is non-arbitrary, that is because any tokening of that sentence is an unmediated presentation of the proposition *snow is white* itself. But if no medium is involved, then *a*

fortiori no symbolic medium is involved: no sentence. So Fodor cannot say that the relation between **snow is white** and *snow is white* is non-arbitrary.

So he must say that it is arbitrary. But if this relationship is arbitrary, then it must be explained why Mentalese sentences are paired off with the propositions in the way they are -- just as it must be explained why 'snow is white' (the English sentence) means *snow is white*: some kind of psychological or historical explanation is called for. (We can't say: 'snow is white' *just* means *snow is white.*) Now the reason why 'snow is white' means *snow is white* is due – at least in part (though perhaps not entirely) – to *what people think*. In other words, it is human cognitive processes that *get* 'snow is white' to mean *snow is white*. In general, in cases where the relation of a symbol to its meaning is arbitrary, it supposed that it is human cognitive activity that makes that symbol have that meaning. So, if the relation between (e.g.) **snow is white** (the Mentalese sentence) and *snow is white* (the proposition) is arbitrary, then the former means the latter because of human cognitive activity. In general, if the relation between Mentalese sentences and their meanings is arbitrary, then it is human cognitive activity – it is *what people think* – that gets the former to mean the latter. But if Mentalese symbols are *given* meaning by what people think, then the meaningfulness of such symbols is *posterior* to human thought, and thus cannot be identical with it.

There is a related difficulty with Fodor's position. Suppose for the moment that there is such a thing as Mentalese. In the quotation above, Fodor said that you (or the relevant sub-personal entity within you) don't apply knowledge of rules to ascertain the meanings of Mentalese expressions; for *y o u* don't represent these rules: you don't *know* what they are. Is this feasible?

Presumably one of the semantic rules of English is that 'red' means *red.* Now suppose that no one knew of this rule. In that case, would it still be a semantic rule of English? What *makes* 'red' mean *red*? Well, semantical rules

are, at least to some extent, *descriptive*: they say what people *understand* by expressions and what they *mean* by them. It would make absolutely no sense to say: 'red' means *red* in English, but no English speaker *takes* 'red' to mean *red* and no English speaker means *red* by 'red'. If no one meant *red* by 'red', and no one took 'red' to mean *red*, then 'red' simply wouldn't mean *red*. So semantical rules are *ultimately* descriptive: they describe how people use expressions -- what they mean by them and how they interpret them. (Again, *primitive* semantical rules are descriptive: it is in this sense that semantical rules are 'ultimately' descriptive. I concede that derived semantical rules – e.g. the rule that assigns *flying penguins are good at three-dimensional chess* to the sentence 'flying penguins are good at three-dimensional chess' – are not descriptive , or at least not *strictly* descriptive. I will elucidate this in a moment.)

It is pretty clear that, if someone is to be able to communicate in English, he must have some grasp of the semantical rules of English: in fact, this is trivially true, as I will now try to show. If someone has no idea what people mean by the word 'red', then any sentence containing that word will, to that extent, be meaningless to him. And if he *does* know what people mean by the word 'red', then *ipso facto* he knows (at least partially) a semantical rule for English: for the semantical rule in question – viz. "'red' means *red*' – just *is* a description (inter alia[237]) of what people mean by that word.

To be sure, not *all* semantical rules are descriptive: the 'derived' ones tend not to be descriptive of actual use. For example it is a semantical rule of English that 'flying penguins are good at three-dimensional chess' means *flying penguins are good at three-dimensional chess*. But this rule does not describe what people *actually* mean, or understand, by the sentence in question; for that sentence has probably never been used. But any *basic* semantical rule – any rule not derived from other semantical rules – *does* describe what people mean

and understand by expressions. The rule "'red' means *red*' is true precisely because people *mean*, and *understand*, *red* by 'red'. (The rule is a somewhat idealized description of the facts, but it is a description no less.) So semantical rules are *ultimately* descriptive.

Given this, it follows trivially that one cannot know a language without knowing the semantical rules for it. For the semantical rules are just descriptions of what people mean by expressions of that language, and of how they interpret them; and if someone doesn't know what people mean by expressions, or how they interpret them, then he simply doesn't know the language. To know the language just *i s* to know what people mean, and understand, by a certain class of expressions. So it is trivial to say that to know a language involves knowing its semantical rules.

Now if there is such a thing as Mentalese, then there is some set of propositions describing the way in which Mentalese expressions are paired off with meanings; i.e. there are semantical rules for Mentalese. (This is trivially true: it would be meaningless, or nearly meaningless, to say of some language L that there were no propositions stating what meaning are borne by expressions of L.) Of course, Mentalese sentence-tokens (supposing that there are such things) are presumably understood. (If Mentalese is the language of thought, then surely *something* in each of us understands Mentalese sentences.) To understand a sentence token is to know what meaning it bears; and what meaning it bears is, ultimately, determined by what *is* meant by it or, at any rate, by what *is* meant – by what thought is intended to be conveyed - by the primitive expressions occurring in it. To know this, in turn, is to know the semantical rule giving meaning to that expression. So with regard to the idea that one (or something with one) could understand Mentalese sentences without knowing the semantical rules for such sentences, that idea is self-contradictory.

Let me elucidate this last argument. The primitive semantical rules of a language L are really just descriptions (if somewhat idealized ones) of the way the primitive expressions of L are used: of what is meant by them and what is understood by them. (It would be meaningless to say that 'red' means *red* but that no one ever understands or means *red* by 'red': for 'red' to mean *red* just *is* for people to mean and understand *red* by 'red'.) Ultimately, what 'red' means *objectively* is determined by what *people* mean by it So there is no way to distinguish knowing what is meant by expressions of a language from knowing the rules that give meaning to those expressions. So with regard to the idea that one could know a language like English without knowing the rules giving meaning to it – that idea is non-sensical.

This point can be brought to bear on the notion that there is a language of thought. The primitive semantical rules for Mentalese (supposing, for the moment, that there is such a thing) cannot be anything other than descriptions of what contents are *in fact* borne by primitive expressions of Mentalese. Suppose that the Mentalese expression **Red** always *in fact* bore the meaning *green*. In that case **red** would mean *green*; and it would be meaningless to say that **red** *actually* meant *red*. In general, if a mental symbol **x** *in fact* bears the message *x*, then it would be meaningless to say that **x** *really* means something other than *x*; it would be meaningless to say that the semantical rule for **x** actually assigns it the meaning *y*. So where Mentalese is concerned, no less than where natural language is concerned, semantical rules are ultimately *descriptive*: there is no distinguishing what is *in fact* meant and understood by **x**, on the one hand, and what it 'really' means – what meaning it is assigned by the semantical rules for Mentalese – on the other. So, where Mentalese is concerned, knowing the semantical rules for Mentalese *ultimately* coincides with knowing what Mentalese expressions means. So with regard to the idea that one could know the rules of Mentalese without knowing the semantical

rules for it, that idea is nonsense.

In conclusion, Fodor's way of blocking the regress discussed above fails. Fodor's way of doing this, recall, is to say that we can understand Mentalese sentences without knowing the semantical rules for them. But we've just seen that this idea is, to be quite blunt, self-contradictory.

I would like to close with one other point. It has been said (though more often merely insinuated) that only a creature that knows a language can have thoughts, can grasp propositions. Given what we've seen, it is clear that this is not the case. It is not possible to know a language L without knowing the rules giving meaning to L. Those rules, for reasons we've seen, must ultimately be grasped in some non-linguistic form. So if a creature is to know a language, it must be able to grasp some propositions in some *non*-linguistic way.

Notes

NO SUCH THING AS MENTALESE REVISITED

Abstract: Another Argument Against the Thesis That There is a Language of Thought.

Jerry Fodor maintains that we think *in* a language – that thought is 'mediated' by some kind of internal symbolism. Every language has minimal units of significance whose meanings cannot be a function of the meanings of other expressions. Therefore, if we think in a language, then we have minimal concepts that are not to any degree 'interdefinable', i.e. that are such that possession of any one of them doesn't presuppose possession of any other concepts. But it is demonstrably the case that there are no such minimal or 'atomic' concepts. Also, if there is a language of thought then any constitutive connections that obtain among complex thoughts become *containment* relations holding among symbols: my concept of 1,423 becomes a symbol that *contains* a concept of 1,422, and *that* concept is itself a symbol that contains my concept of 1,421, and so on. So my concept of 1,423 contains each of these other 1,422 concepts in the same sense that the word 'hereditary' contains the word 'red'. A consequence of this is that, when I exercise my 1,423-concept I am *ipso facto* exercizing my 1,422-concept, my 1,421-concept, and so on; that is, when I think about 1,423, I am *ipso facto* thinking about 1,422, 1,421, and so on. But this is decidedly implausible. There are two possible responses to this: one is to deny that constitutive relations among concepts are containment relations; the other is to deny that there are any constitutive relations among concepts. Fodor takes the latter path – so, on his view, I can have a concept of 1,423 without having a concept of 1,422 or of 1,421. But such an extreme representational 'atomism' is demonstrably false; and as long as constitutive relations among concepts

are not crudely construed as containment relations, there is no reason to advocate such a position. Indeed Wittgenstein observed long ago that constitutive connections among concepts are not to be equated with relations of decomposability or containment: one's concept of azure is constitutively related to one's concept of navy blue, but neither concept is decomposable, even partially, into the other.

I. According to Jerry Fodor, one grasps the proposition *Fred is coming home* by way of grasping some *sentence* (or, at any rate, some kind of symbol) that has this proposition as its meaning. Consequently propositional attitudes are mediated through *sentences* that express the propositions in question. For me to hope that Fred is coming home is for me to stand in a certain relation to a sentence that means *Fred is coming home*; for me to fear that Fred is coming home is for me to stand in some other relation to this same sentence; and so on *mutatis mutandis* for every other propositional attitude that I might have towards *Fred is coming home.* Finally, for me to make some kind of an *inference* involving this proposition is for me to recognize that the aforementioned *sentence* stands in some relationship to some other *sentence*; for example, for me to infer that *someone or other is coming home* from the proposition *Fred is coming home* is for me to recognize that a sentence meaning *someone or other is coming home* stands in a certain relation (a computational – that is, a syntactico-morphological – relation) to a sentence meaning *Fred is coming home.*

Basically, one grasps propositions entirely *through* sentences that have those propositions as their meanings: consequently, there is a 'language of thought'; cognitive (i.e. proposition-bearing) mental activity is thus 'symbolic'. In a word: we think *in* a language.

One point must be made clear. In maintaining that we think *in* a language,

Fodor does *not* mean that English speakers think *in* English, that Spanish speakers think *in* Spanish, and so on. He means that each of us thinks in an innate, internal language. (Fodor holds – rightly, in my view – that one must be able to think *in order* to learn a language like English or Spanish. So he is committed to holding that the language of thought is not a learned language like English or Spanish.)

Why does Fodor think that there is a language of thought – that we think *in* language? His reasoning is basically this. We are computers. (In any case, the 'best models of cognitive activity are computational'.[238] So, pending some extraordinary development in cognitive psychology, we should work on the assumption that we are computers.) A computer is an information processing engine. Information is embodied in a computer in a kind of internal code: in something that is, in effect, a language. Consequently, a computer processes information by processing strings of code – things that are, in effect, sentences. So any access to information, and thus to propositions, that a computer has is mediated through its access to a sentence that bears that information.

If this characterization of what it is to be a computer is correct, and *if* we are computers, then it very much seems to follow that *our* mental activity consists in sentence processing, and that *our* access to information is mediated entirely through sentences that bear that information.

In this paper I will set forth one reason – not found, to my knowledge, in the growing literature on this question – to reject the idea that there is a language of thought. The argument set forth in this paper, supposing it cogent, has a fair amount of scope. As we just noted, *if* we are computers, then there must be a language of thought. In any case, this follows *if* the just given characterization of what it is to be a computer is accurate: and that characterization is a fairly standard one. So *if* there is no language of thought,

then (assuming the correctness of that characterization), it follows that we are not computers: and this, of course, would not be a trivial result.

It must be said that virtually everyone who believes in the computational theory of mind believes that there is some kind of language of thought or, at any rate, some kind of internal code. And *prima facie* the former position seems to presuppose the latter. In any case, I think that Fodor is right to say that *if* we are computers, then there is a language of thought.

But of course the question whether there is a language of thought is worth investigating in its own right, regardless of any bearing it might or might not have on the validity of this or that theory of mind.

The question whether there is, or even could be, a language of thought is an acutely controversial one. Usually, when the idea of a language of thought is attacked, the attack is to the effect that the intentionality borne by symbols is derivative of that borne by mental entities: basically, symbolic intentionality – and thus the very existence of symbols – *presupposes* some other kind of intentionality and, therefore, some non-symbolic mental activity.[239] Those who believe that there is a language of thought counter-attack by saying that this criticism begs the question: *if* there is a language of thought, *then* with regard to the age-old view that symbolic intentionality is, in general, derivative, that view is simply false and needs to be thrown out or circumscribed.[240]

The argument set forth in this paper represents a completely different line of attack and is therefore not vulnerable to anything that has been said in connection with the question whether symbolic intentionality is derivative.

II. Let us refer to one's ability to represent an object or property as a 'representational capacity' or as a 'concept'. (So I will use the terms 'concept' and 'representational capacity' interchangeably.[241]) So, for

example, in virtue of being able have thoughts about Bob, one has a certain representational capacity or concept. In virtue of being able to see objects *as* being of a certain color, or *as* having a certain shape, one has a certain representational capacity.

Now these days it is widely held that representational capacities have 'constitutive connections' to one another; that they come in *bundles* and not singly. To embody this point in some terminology: some kind of 'molecularism' or 'holism' – some kind of 'non-atomism' -- about representational capacities is widely countenanced; and representational 'atomism' is generally held to be false. Significantly, Fodor himself is an atomist. [242]

What does all of this mean? What does it mean to say that representational capacities have 'constitutive' connections to one another? Obviously we can see physical objects *as* being triangular. Presumably, one could not represent things as being triangular to oneself unless one could also represent various other things to oneself: e.g. space, straight lines, angles. So if one can represent things to oneself as being triangular, then necessarily one can represent things to oneself as being straight, and as being in space, and as intersecting. Also, one must presumably have some grasp of number: to grasp a triangle, one must grasp the idea of *three* different lines. And to grasp the concept *three*, one must surely grasp other number concepts. (One couldn't grasp the concept *three* unless one grasped the concepts *two* and *one*, not to mention various other concepts.) The concept of number would seem to have constitutive connections to the concept of order: if one grasps the former concept, surely one *necessarily* has some concept, however vague, of the latter.

The 'necessarily' here is not *causal*. The connection between being able to represent things to oneself as being in space does not have a *causal*

connection to one's being able to represent things to oneself as triangular. The connection is much stronger than that: *what it is* to be able to represent things as being triangular is to be able to see them as being in space, and as being bounded by intersecting lines, etc. So the one representational capacity *makes up* the other.

Prima facie every representational capacity one might think of has constitutive connections to others. This is obviously true of those capacities whereby we are able to represent *complex* things (in their complexity) to ourselves. And it is also true – though, perhaps, less obviously – of our ability to represent to ourselves those things that are 'simple' or, at any rate, those things that are the best candidates for that title. In fact, by a curious reversal, it sometimes takes a *heavier* representational arsenal to represent simple things (in their simplicity) to oneself than it does to represent complex things. Just as a formidable technology is needed to isolate, say, pure oxygen or to produce a hard vacuum, so a formidable cognitive machinery is needed to isolate, say, the concept of a point or that of a line. Surely a space-time point is as good a candidate as any for the title 'simple object'. But the concept of a space-time point is a highly sophisticated and derivative concept, and is necessarily deeply embedded in other concepts; for the concept of a point can (at least by some reasonable accounts) only be grasped *via* other concepts. A point is the intersection of two lines or is a kind of degenerate line; or a point is (this is Whitehead's definition) a maximal class of spheres such that, given any two of them, one contains the other. (Whitehead's definition is counter-intuitive but logically unexceptionable.) And one's ability to represent *these* concepts, in their turn, is necessarily embedded in yet other representational capacities. (A line can only be thought of in a roundabout way: as, for example, the intersection of two planes or as a degenerate space. So one's ability to represent lines is necessarily embedded in some other

representational capacities – e.g. one's ability to represent planes etc.) Basically, *ontological* simplicity often corresponds to staggering conceptual and, therewith, cognitive complexity.

What these all too brief remarks are meant to illustrate is that there is good reason to believe that representational capacities are constitutively (not just causally) connected to one another; that they are not *discrete* -- not separable into distinct, mutually exclusive bundles.

I will provide a genuine argument for this thesis later on. So far, I've only been trying to show that a non-atomism about representational capacities is to be given some credence.

Let me now outline the argument I will set forth. Any language that is represented in our minds – and presumably the language of thought, supposing that it exists, is such a language – must have a finite number of primitive symbols. These symbols will, of course, be completely discrete; they will be completely separate from each other. If two symbols are not separate, that means one *contains* the other. And if one contains the other, then *ipso facto* one of them is not primitive. (The word 'doghouse' contains the word 'house' and is *ipso facto* not a primitive expression of English.) So the primitive expressions of the language of thought must be separate – must not contain one another. Now if we think in a language, then these expressions will be our *basic* concepts: if we think *in* a language, then expressions are our concepts, and primitive expressions are our primitive concepts – our *basic* representational capacities. But if the conceptual non-atomism just outlined is correct – and I will argue at length (in section VI) that it is -- then there are no primitive concepts: there are no *basic* representational capacities: any such capacity necessarily exists in a 'space' of others. Therefore, we cannot literally think *in* a language: there is no language of thought – there is no internal code that mediates our thoughts.

Now I would like to delineate and defend this line of thought.

III. According to the thesis under examination, we access propositions – e.g. *a square peg of such and such dimensions cannot fit in a round hole of thus and such dimensions* – by way of grasping a sentence that expresses that proposition. It would very much seem to follow that we grasp the concepts composing this proposition by way of *symbols* that represent them. *If* I can access the proposition *a square peg of such and such dimensions cannot fit in a round hole of thus and such dimensions* only by way of accessing some sentence that has this proposition as its meaning, then, surely, I can access the concepts composing this proposition by way of grasping certain symbols that denote them: a symbol meaning *round*, one meaning *square*, etc.

So, assuming that there is a language of thought, one grasps the concept *roundness* wholly *by way of* grasping a *symbol* that denotes (or expresses) this concept. Supposing that **roundness** is the Mentalese symbol denoting the concept *roundness* -- one's ability to represent roundness to oneself is, in effect, *identical* with the Mentalese symbol **roundness** (or, to speak more strictly, with one's being in possession of this symbol): if we think *in* Mentalese, then the Mentalese expression **roundness** is one's concept of roundness. In general, if we think *in* Mentalese, then Mentalese expressions are our concepts.

As we noted earlier, it is widely held that representational capacities are not *discrete*; they permeate one another; or, to use a different image, they come only in mutually entangled bundles. But *symbols* are utterly discrete. To make this clear: one's ability to represent triangles to oneself is permeated by one's ability to represent lines, space, angles, etc. to oneself. And each of these abilities is, by all accounts, permeated by other representational abilities. But any *symbol* that denotes the property of being a triangle or a line

is perfectly discrete; the words ‘line’, ‘angle’, ‘plane’, ‘point’, ‘rigidity’, etc. are as discrete – as distinct from one another – as they could possibly be. Symbols are, of their very nature, discrete things; it makes no sense to speak of *symbols* that have ‘constitutive connections’ to one another or that ‘permeate one another’. Such talk applies to concepts (and abilities), but not to symbols. Therefore such talk does not apply to Mentalese *symbols*, supposing that there are such things. The Mentalese symbols ‘line’ and ‘triangle’ are (if they exist) as distinct from each other as the English words ‘line’ and ‘triangle’.

So *if* representational capacities are interconnected, and do not come in mutually isolable packages, then these capacities are in no way identical with discrete symbols; and, consequently, my ability to think about roundness is in no way identical with my being in possession of a Mentalese word that means *round.*

This point applies to our ability to grasp propositions. Obviously *if* one’ s ability to think about triangles is constitutively bound up with one’s ability to think about e.g. planes, then so is one’s ability to grasp propositions involving the concept of triangularity. Indeed, it is highly reasonable to suppose – and deeply counter-intuitive not to suppose – that one’s ability to grasp the proposition *alpha is a triangle* is constitutively bound up with one’s ability to grasp the propositions *alpha is planar* and *alpha is in space* etc. Basically, not surprisingly, what is true of our abilities to represent concepts is true of our abilities to grasp propositions: the latter, like the former, are interconnected.

Of course, according to the thesis under scrutiny, one’s grasp of the propositions *alpha is a triangle* and *alpha is planar* is identical with one’s grasp of *sentences* that express these propositions. As we just noted, one’s ability to grasp the one proposition is not isolable from one’s ability to grasp

the other. But the *sentence* ‘alpha is a triangle’ is utterly distinct from the *sentence* ‘alpha is flat’; and the same, presumably, is true of the Mentalese translations of these sentences. So the Mentalese sentences **Alpha is flat** and **Alpha is a triangle** are distinct; and, for reasons already discussed, if we think in sentences, then one’s ability to grasp the corresponding propositions is identical with one’s being in possession of tokens of these sentences. Obviously the *sentences* are completely distinct; they are as distinct as can be. So, presumably, one’s possession of the one sentence is as distinct as can be from one’s possession of the other. But obviously one’s ability to grasp the one proposition is *not* isolable – at least not in its entirety – from one’s ability to grasp the other. So – if some kind of representational non-atomism is the case – then one’s ability to think *alpha is flat* is *not* identical with some sentence.

To summarize: we do not think *alpha is flat* ‘in’ a sentence; we do not grasp that proposition by way of grasping (or tokening) an internal sentence that expresses that proposition. For any *sentence* that means *alpha is flat* will be completely and utterly distinct from any sentence that means *alpha is a triangle*. But one’s grasp of the one proposition is very much *not* separate from – is indeed essentially connected to – one’s ability to grasp the other. So we don’t grasp e.g. *alpha is flat* ‘in’ any language (in any case, we do not *have* to so grasp it); we do not – at least not in general -- grasp propositions ‘in’ any language: there is no language of thought.

The foregoing was really just a sketch and, like any sketch for a controversial position, is open to innumerable objections. Before considering some of these objections, I’d like to slightly touch up what was just said. One point in particular needs to be addressed.

Obviously the English words ‘line’ and ‘triangle’ are totally different from each other. But the corresponding Mentalese words are not different

from each other. Obviously one's concept of triangularity is bound up with one's concept of linearity. This is perfectly consistent with the thesis that we think in Mentalese. We just say that the Mentalese expression for triangle is defined in terms of the Mentalese word for line. In English 'triangle' is a simple expression. In Mentalese – since this is the medium of thought – the word for 'triangle' is complex, and its connectedness to other concepts -- e.g. 'line', 'plane' -- can be read off its formal structure: so the Mentalese expression for 'triangle' would be (perhaps) 'closed, three sided, planar figure.' Given this, it isn't true that Mentalese expression for 'triangle' is distinct from the Mentalese symbol for e.g. 'line': the two symbols are interconnected in just the same way as the corresponding concepts. So your argument falls through.

This is an important counter-objection. But there are several reasons why it doesn't work. First of all, any language has 'minimal units of significance' (morphemes). This is not a contingent fact about natural languages; it is a necessary fact about any language – at least any that can be known by a human. Mentalese, supposing that it exists, can obviously be known. So Mentalese cannot have an infinitely large primitive vocabulary. (Surely we should reject any model of cognition that imputes to us knowledge of *infinitely* many primitive semantic rules.)

If we think *in* Mentalese, then the morphemes of Mentalese are the basic units of our thought. Those concepts would be *primitive*, at least from our point of view. This means that they would not have constitutive interconnections to one another: if one concept C1 is to any degree *constitutive* of some other concept C2, then *ipso facto* C2 is not primitive; it represents some kind of an intersection or derivative of other concepts.

So if we really think *in* a language, then some of our concepts are

primitive: for a language has minimal units of significance. Consequently, if we think *in* Mentalese, then its primitive units of significance – and it must have such units -- are our primitive concepts. So we must have primitive concepts. Primitive concepts are *isolated*; given that they are *primitive*, they *ipso facto* do not presuppose other concepts and thus exist in a kind of vacuum.

But there are no primitive concepts. So we do not think *in* a language. Of course, the view that there are no primitive concepts needs to be defended: I will provide an extensive defense of it in a moment. This defense is not, so far as I know, already in the literature and, importantly, it is therefore not considered by Fodor in *Concepts*.

Permit me to clarify the line of thought just set forth. The term 'triangle' can be defined in terms of 'line', 'intersection' etc. Given this, it seems feasible to maintain that the Mentalese expression **triangle** is defined in terms of the Mentalese expressions **angle**, **intersection**, etc. As we've noted, Mentalese, like any other (knowable) language, must have minimal units of significance. So the definition for **triangle** must terminate in such units. Obviously, these minimal units are not defined in terms of one another: they are wholly discrete. So once again consider the Mentalese term **triangle** and suppose that m1, m2…mn are the ultimate Mentalese expressions in terms of which **triangle** is defined. In that case, each of m1, m2…mn is totally isolated from every other expression in the language; each of these terms corresponds to a concept that does not presuppose any others. So we cannot say about m1, m2…mn what the object says about 'triangle' and 'line'. It is arguable that, while the *English* expression 'triangle' is distinct form the *English* expression 'line', the same is not true of the corresponding Mentalese expressions. But we cannot coherently take this view of m1, m2…mn: these expressions are distinct and completely non-interdefinable. Consequently,

supposing we think *in* Mentalese, these expressions are the minimal units of thought, and they bound our thought: they are, in short, concepts that have no constitutive connections to one another.

But *there are no such concepts.* One of the most striking and celebrated failings of logical positivism was precisely its belief that there are such concepts: that there are primitive, foundational concepts such that all concepts are either identical with, or built out of, these.

So our initial point prevails. *Concepts* come bundled; *lexical* items do not come bundled. Therefore, lexical items cannot be our concepts. Therefore we do not think *in* lexical items; we do not think *in* any language.

There is another reason why the objector's point fails; it conflates semantic analysis with conceptual analysis. It makes thoughts like *that is a sphere* be identical with very different, but necessarily equivalent, thoughts – e.g. thoughts like *that is a closed three-dimensional figure of uniform curvature.*

Consider the sentences 'alpha is an object whose surface points are a given distance from some particular point' and 'alpha is a closed three-dimensional figure of uniform curvature.' Obviously these sentences necessarily have the same truth-conditions. But, just as obviously, they are not *synonymous.* The propositions they express involve very different concepts. (One involves the concept *curvature,* the other does not; one involves the concept *point,* the other does not.) So, plainly, these propositions are different from each other, even though they are necessarily equivalent.

Now *each* of the two propositions just considered gives an unexceptionable *analysis* of the sentence 'that is a sphere.' Each of these propositions makes explicit what is, in some sense, implicit in 'that is a sphere.'

At the same time, the *sentence* 'alpha is a sphere' is not, on that account,

ambiguous. It is a sentence that has *one*semantics, but that admits of different *conceptual* analyses. In fact, there are as many different *conceptual* analyses of ‘that is a sphere’ as there are axiomatizations of geometry: infinitely many. But ‘that is a sphere’ is surely not *infinitely* ambiguous; it is not ambiguous at all. So conceptual analysis and semantic analysis are not the same thing: a single semantic analysis may correspond to a multiplicity of conceptual analyses.

What is the bearing of this on the question whether we think *in* a language or symbolic code? According to the proposal under examination, we grasp the proposition *that is a sphere* entirely *through* some Mentalese sentence that has that proposition for its meaning. Now our ability to grasp that proposition obviously has constitutive connections to our ability to grasp other propositions (e.g. *that is planar*). So given that we grasp these propositions wholly *through* sentences, it follows that the constitutive connections holding between these different propositional capacities must be reflected in the structure of the sentences themselves: the Mentalese sentence for *that is a sphere* must quite literally *contain* the Mentalese sentence for *that is planar* and also for *that is in space* and also for *that has a continuously changing perimeter* and *that has a uniform radius.* For the only *constitutive* relations that obtain among symbols are *containment* relations: like the relation holding between the expression ‘figure with three sides’ and the expression ‘sides’. So, if we think in Mentalese, then the thought *that is a sphere* must, in a literal sense, ‘decompose’ into some thought containing *all* the concepts that are constitutively connected to that thought: it must decompose into a thought that is composed of the concepts for curvature, closure, space, uniformity, and so on. Basically, if we think in Mentalese, then the thought *that is a sphere* must be identical with some *analysis* of that thought: with some thought like *that is a closed three-dimensional figure of*

uniform curvature or *that is a closed three-dimensional figure with a uniform radius*.

But the thought *that is a sphere* is *different* from, though constitutively connected with, the thought *that is a closed planar figure of uniform curvature.* If they weren't different, then the thought *if that is a sphere, then that is a closed planar figure of uniform curvature* would be as trivial, and would involve as little intelligence, as the thought *if that is a sphere, then that is a sphere*. So the thought 'that is a sphere' – while *constitutively* connected with the ability to think *that is a closed planar figure of uniform curvature* – does not in any sense *contain* the latter thought or *vice versa.* Now if we think in symbols, then constitutive connections among thoughts are containment relations among symbols. So the thought *that is a sphere* would literally contain the thought *that is a closed figure of uniform curvature* and also *that is a closed figure of uniform radius* – along with every other thought constitutively connected to *that is a sphere.* But this picture doesn't correspond to how our thoughts interrelate: the thought *that is a sphere* does not – in any but an extremely tenuous and metaphorical sense – contain the thought *that is a closed figure of uniform radius* or, indeed, any of the other thoughts that are constitutively connected to it.

This, I think, is one of the great weaknesses of the 'symbolic conception of thought': it turns all constitutive relations among thoughts into containment relations. Further delineation of this idea shows just how untenable it is.

V. Given what we've just observed, there is yet another reason to hold that we don't think *in* Mentalese. As we noted a moment ago, the only way that a *symbol* can be constitutively connected to another symbol is if one of them *contains* the other – literally *constitutes* it (as e.g. the symbol 'figure with three sides' contains the symbol 'sides'). So if we think *in* Mentalese

sentences, then the constitutive relations obtaining among our thoughts become *containment* relations holding among sentences. To illustrate this: Obviously our ability to think *that is a sphere* presupposes our ability to think *that is a closed figure.* So if we grasp the proposition *that is a sphere* entirely *through* some sentence having that proposition for its meaning, that sentence must actually *contain* some sentence meaning *that is a closed figure.*

But constitutive relations among concepts are demonstrably not *containment* relations. This is most obvious in the case of color thoughts. Suppose you have a concept of some *specific* shade of pink – say pink1. It is impossible that you should be able to have *this* concept unless you had concepts of other colors or shades. (I will give an extensive defense of this in a moment. But it obviously has plausibility.) Any given color-concept necessarily exists in a space of other color concepts. But one's concept of pink1 is not some complex concept consisting of one's concepts of other shades of pink or other colors. Such a proposal would make no sense. First of all, the idea that one's pink1 concept is *composed* of other color-concepts that one has – say, concepts C1, C2…Cn --- presupposes that colors can be *defined* in terms of other colors. But this is patently not the case; colors are famously *not* inter-definable. (It may be true that red and white make pink; but 'pink' cannot be defined in terms of 'white' and 'red'; and the concept *pink* cannot be analyzed in terms of the concepts *red* and *white.*) [243]

Let us sum up: according to the proposal in question, one's pink1-concept is some symbol that *contains* all the concepts constitutively connected to one's pink1-concept; and these other concepts are themselves symbols. So one's pink1-concept, according to the proposal in question, becomes a complex symbol defined in terms of – indeed, composed of – other color-symbols. But color-symbols are famously not mutually-definable; and color concepts are famously not decomposable into other color concepts. 'Azure'

is not definable in terms of ‘blue’; and the concept *azure* is not analyzable in terms of the concept *blue.* (Of course, one can say ‘azure is a certain shade of blue’; but this is no analysis of the concept *azure.* The concept *azure* is not some compound consisting of the concept *blue* and some other concept.) But one couldn’t have azure-thoughts unless one had blue-thoughts; these are constitutively interconnected concepts. So the lesson is that constitutive connections among concepts are not containment relations.

There is a second reason to think that one’s color-concepts do not in any literal sense *contain* other color concepts. *Any* color-concept exists in a space of other color-concepts, i.e. any color-concept has constitutive connections to other color concepts. So if constitutive connections among concepts were containment relations among symbols, then any color concept one had would be some color symbol, and that color symbol consists of other color concepts, and each of *these* color concepts would, in their turn, be color-symbols. But then each of these symbols would, by the same reasoning, consist of yet other color-symbols; and each of these symbols of yet other color symbols, and so on *ad infinitum.* Basically, any color thought would actually *consist of* (not presuppose: but actually be built out of) other color-thoughts. So in having a single color thought, one would actually be having infinitely many color thoughts: but this is not the case; indeed, it would be absurd to suppose that it were the case.

There is a way to block this infinite regress: say that each color-concept C1 is defined in terms of other concepts, and these latter in terms of still other ones, but that sooner or later we loop back to C1, i.e. we reach some color concept that is defined in terms of C1; so that only *finitely* many color-thoughts are implicated in any one color thought. But this is patently just as untenable as the regress mentioned before; it makes it a *pre*condition for being able to have e.g. azure-thoughts that one *already* be able to have azure-

thoughts: one must be able to have azure thoughts *before* one is able to have azure thoughts.

So constitutive connections among concepts are not containment relations of any kind, and *a fortiori* are not containment relations among symbols. We've seen one illustration of this point. But another one is probably in order.

Presumably, to have the concept *1,143*, one must have the concepts *1,142* and also the concept *1,141* and so on. Now if constitutive connections among concepts were identical with containment relations among Mentalese symbols, then one's *1,143* thoughts would be 1,143-symbols that *consisted* of one's 1,142-symbol and one's 1,141-symbol, and so on. But it is very hard to believe that when you have a thought like *there are 1,143 apples in the warehouse*, you are having a 1,142-thought, and a 1,141-thought, and so on. At the same time, it is patently obvious that someone who was totally oblivious to the concept *1,142* would *ipso facto* be incapable of grasping the concept 1,143; it is patently obvious, in other words, that the latter concept is constitutively connected to the former. So given (i) that these concepts are constitutively connected, and given (ii) that 1,143 thoughts are not cobbled together out of 1,142-thoughts and 1,141-thoughts and 1,140-thoughts…and so on, it follows that *constitutive* connections are not identical with any kind of *containment* relation and *a fortiori* are not identical with containment relations holding among symbols.

This consideration demolishes one of Fodor's main arguments for atomism. Fodor says that *there are* primitive concepts – indeed, that *most* of our concepts are primitive. By this, Fodor means, or at least implies, that we *do* have a stock of concepts C1, C2…Cn such that each of these concepts lacks *any* constitutive connections to others. (Fodor obviously grants that we can *construct* concepts that are constitutively connected to others; e.g. the

concept *is either a man or a horse* is obviously constitutively connected to the concepts *man* and *horse.* And, Fodor also admits, concepts belonging to '*bona fide* axiomatic systems' – e.g. triangle, square – have constitutive connections to others: the concept *triangle* can, Fodor obviously admits, be defined in terms of the concepts *three-sided closed figure* and therefore has constitutive connections to these others. And Fodor is willing to grant that, besides these, there may be a few other cases of constitutive interconnections among concepts: e.g. the concept *bachelor* is constitutively connected to those of *male* and *married.* But he thinks that the concepts that populate our thoughts are, for the most part, *not* constitutively connected to others: the concepts *doorknob*, *horse*, *man*, *carburetor* – practically any concept entering into a thought exists in isolation from all others. One of Fodor's main arguments for this claim is that most of these concepts are not *definable* in terms of others. The concept *man* is not *definable* – not *decomposable* into other, simpler concepts. Concomitantly, there is no *complex* symbol S such that S is necessarily coextensive with 'man'.

Fodor is, in my view, right to say that most concepts resist *decomposition* into others; in other words, Fodor is right to maintain that concepts are not typically *composed of* – do not typically *contain* – other concepts. (Fodor makes a good case for this.) But this doesn't mean that concepts do not have constitutive connections to one another: for *constitutive relations among concepts are not containment relations,* as we've just seen. So this undercuts Fodor's primary argument for atomism.

VI. So far I've taken it for granted that there are *no* concepts C1, C2…Cn such that each of these has no constitutive connections to others and such that all our concepts are identical with, or decomposable into, these. But, obviously, this assumption of mine is contentious: now I will discharge it. I will present an argument for 'non-atomism' that is completely general, i.e. it

shows that *any* concept has constitutive connections to others, not just some concepts belonging to some privileged class. (Sellars' classic argument against atomism does not concern, at least not explicitly, *all* concepts. It concerns, at most, only empirical concepts, i.e. concepts implicated in sense-perception.[244]) My argument has, so far as I know, never appeared in the literature.

One cannot *just* see an object: one must see it *as* having this or that property; as being round or square or blue. And one cannot *just* think about, or imagine, an object. One must think about it *as* round or square or sweet. In general, one cannot *just* represent an object to oneself; one must represent it as having these or those properties. Now it seems that if one can represent one property to oneself, then one can ipso facto represent others to oneself: each representational capacity has 'constitutive' connections to other representational capacities. For example, consider the property of triangularity. If one can represent something to oneself as triangular – if one can see or think about or imagine…something as triangular – then *ipso facto* one can represent things as having other properties: as e.g. being straight, being planar, being closed. And in order to grasp any of these properties, one must be able to grasp others. For example, if one can think about the property of being closed, one can think about the property of geometrical continuity (for a closed figure is one with a continuous periphery); and if one can think about this property, then presumably one can think about others; for the property of continuity is presumably not a simple one. (Exactly what one must grasp to grasp the property of continuity is a delicate question. But it seems unlikely that one could grasp this property in a vacuum. Presumably one must have some conception of order and, presumably, of location in some kind of space. And, by all appearances, these concepts are, in their turn, not simple ones.) To sum up, it is pretty clear that one could not represent the

property of triangularity to oneself unless one had various other representational capacities.

I submit that this point can be generalized to *all* properties: given *any* property P, if one is to be able to represent things as having property P, then one must be able to represent things to oneself as having properties other than P. How is this to be shown? First of all, all properties are individuated, at least in part, by their relations to other properties. And it is, as I will now show, impossible that one should be able to grasp one property (i.e. to represent things to oneself as having that property) without having any conception at all of any other property.

It is an essential fact about the property of weighing 300 lbs that, if something has that weight, it weighs more than 200 lbs and less than 1000 lbs. It is an essential fact about the property of being square that if something has that property, it has the property of being rectangular and does not have the property of being round. It is a essential fact about the property of being pink that, if something is pink, it is closer in color to something which is red than to something which is green. These points hold *mutatis mutandis* for any property one might think of. Each property exists in a 'space' of other properties; and, I suggest, one can no more grasp a given property without being able to locate it in a space of other properties than one could have the concept of a given spatial co-ordinate without being able to relate it to other co-ordinates.

This is *obviously* the case with degree-properties: properties like that of weighing a thousand pounds or even the property of simply being heavy. For such properties are clearly relational; and the information one acquires in learning of a thing that it has such a property has a *patently* contrastive or relational character. (Surely one couldn't grasp the property of weighing 1000 lbs unless one could grasp the properties of weighing more than 1000

lbs and of weighing less than 1000 lbs. And – what is connected with this -- the information that one obtains in learning of a thing that it weighs 1000 lbs is obviously contrastive: in learning a thing a thing weighs 1000 lbs one is learning, inter alia, that it weighs *more* than one pound.) The question is whether this point (mutatis mutandis) holds of *all* properties. For example, is the property of being a certain shade of pink in any sense a 'degree'-property? Is it, in any relevant way, like the property of weighing 1000 lbs? Or could one grasp this property without grasping any other properties? I now wish to show that what is obviously true of one's grasp of degree properties (and of *obviously* relational properties) is true of one's grasp of *any* property, even those that seem quite non-relational.

For the sake of argument, suppose that there were a creature that could represent to itself the property of being a certain shade of pink but that could not represent a thing as having any other shade. (Let P be this shade.) To that creature, it would literally be inconceivable that an object should be of a color other than P. So from that creature's view-point there would be no distinction between having the property of being P and that of simply being colored. So this creature wouldn't really grasp the property of being *P* (of being a specific shade of *pink*) after all. To be P is to be a specific shade of a specific color. If you are completely incapable of representing that specificity, then you can't represent pink, let alone some specific shade of pink. Obviously a creature that can represent only *one* color can't distinguish *between* colors. And a creature that can't distinguish between colors is incapable of representing the *specific* character of any color; and thus cannot represent any particular shade of pink. So with regard to the supposition that a creature could represent just *one* shade to itself, that supposition is a self-defeating one.

To sum up, if a creature can't represent multiple colors to itself, then it can't grasp what is specific about any a particular color. And if a creature

can't do the latter, then it obviously can't represent that color to itself at all. So a creature must be able to represent multiple colors to itself if it can represent any.

Before continuing, I'd like to present a different way of establishing this same point. Once again, for the sake of argument suppose that some creature could represent to itself only some specific shade of red R. Since this creature, by hypothesis, wouldn't be capable of considering *different* shades/colors, and wouldn't be able to distinguish between *different* colors, its concept of color would coincide with its concept of R. From that creature's standpoint, it would literally be inconceivable that anything colored should not have color R; for to be able to entertain such a concept, the creature would obviously have to conceive of colors *different* from R and this, by hypothesis, is just what it cannot do. But obviously any adequate concept of R – any concept that deserves the name – must be much more specific, must contain much more differential information, than a concept of color in general. So the creature in question *wouldn't* really have any concept of R: it would, at most (though even this is almost certainly false, as I will show presently), have a concept of *color* in general. But it could not be said to have a concept of any *specific* shade/color and thus could not be said to have a concept of R. So, with regard to the idea that a creature might have a concept of a *specific* color without having a concept of any other colors, this supposition is self-defeating.

An exact analogue of this argument can be given for *any* property: the property of square-hood, of being an A-flat, of having the timbre of a French-horn. To take one of these examples: suppose that a creature could represent to itself the pitch *A-flat* but couldn't represent any other pitch. In that case its concept of pitch in general – of having *some* pitch or other – would coincide with its concept of being an *A-flat.* But any concept of A-flat that deserves

the name will be much more specific than a concept of having some pitch or other. So it cannot be maintained that a creature could have a concept of (e.g.) A-flat without being able to distinguish *different* pitches within the genus of pitch in general. I should address an objection:

Maybe you're right to say that a creature couldn't have a concept of some specific color without having a concept of multiple colors. But – by your own admission, practically – a creature could have the concept color *(i.e. the concept of being colored – though not of being e.g. mauve) without having a concept of multiple colors. So doesn't this show that a creature could have a single concept without having a multitude of them?*

No, it does not show this. For what we said about having a concept of some specific color (say, R) applies (mutatis mutandis) to the having of *any* concept, including in particular the concept of being colored (i.e. of having *some* color or other). For the sake of argument, suppose that some creature had the concept of being colored, but didn't have any concepts besides. Then that creature's concept of being colored would coincide with its concept of merely being propertied (i.e. of having *some* property or other), as that creature could not – by hypothesis – consider, and thus distinguish between, a multitude of properties. But any concept of being colored that deserves the name is surely much more specific than a concept of merely being **properties.** So the supposition that a creature could have the concept of being colored without having a multitude of other concepts is incoherent. To sum up, there is good reason to believe that, if a creature is to have a concept of one property, then it must have a concept of several.

There is more to say on this matter. For a creature to be able to represent (e.g.) a specific shade of pink to itself, it is not *just* necessary that it be able to

represent other colors to itself. It would seem that to grasp the *specific* character of (e.g.) P, one must (in some way, at some level) grasp *how* P differs from other colors and shades. To grasp the specific character of P involves grasping what is *distinctive* about it; and this involves grasping *how* it is distinguished from other shades and colors. Just to be clear: Grasping what is distinctive about a color presumably involves what distinguishes that color from other colors; and this involves (maybe is identical with) grasping *how* that color differs from other colors. So being able to represent P to oneself (and to represent things as having P to oneself) involves, not just being able to represent other colors to oneself, but being able to have some kind of a grasp (at some level, in some way) of *how* P relates to these other colors. And of course, exactly this is true of one's grasp of each of these other colors. So being able to represent a single property to oneself would seem to involve grasping how a *multiplicity* of different properties relate to one another; it involves having an interconnected or articulated multiplicity of representational capacities.

This exact argument works for any property, not just P. If you can't represent shapes other than squares to yourself, then you can't distinguish squares from other shapes: for to make such a distinction, you'd have to be able to represent these other shapes to yourself (and, by hypothesis, you cannot represent any shapes besides squares to yourself). If you can't represent these other shapes to yourself, then you can't represent the *specific* character of squares to yourself. And if you can't do that, you can't represent squares to yourself. Further, to represent the *specific* character of square-ness to oneself – i.e. of what is distinctive about squareness: what *distinguishes* it from other shape-properties -- one must have some kind of a grasp (at some level, in some way) of *how* square objects are (geometrically) different from objects that have a different shape. (And, of course, to represent any of these

other shapes to oneself, one must be able to grasp *that* shape's specific character and thus *how* that shape differs from other shapes.) So one can grasp a specific shape-property only if one has some kind of a grasp of a whole 'space' of such properties:one must be in possession of an articulated network of representational capacities.

Conclusion

A language can be seen as a two-tiered structure. It must comprise basic expressions (e.g. 'dog', 'cat') and also complex expressions whose meanings are determined by those of the basic expressions composing them (e.g. 'the dog ate the cat'). Any language -- at least any knowable language -- must comprise basic expressions. If a language contained no basic expressions, then the meaning of any given expression would be a function of the meanings of other expressions, and the meanings of each of *those* expressions would be a function of the meanings of yet further expressions, and so on *ad infinitum.* Consequently, knowledge of the meaning of any expression of that language would presuppose knowledge of an *infinitely* large lexicon; and we may infer, on both logical and psychological grounds, that such a language would not be knowable.

If we think in a language, then the language of thought must have a similarly two-tiered structure. It must comprise basic expressions, whose meanings are not a function of the meanings of other expressions, and complex expressions, whose meanings *are* a function of the meanings of other expressions. The basic expressions would be the minimal units of our thought; these basic expressions would be *basic* concepts or representational capacities. Given this, let C be an arbitrary one of these basic concepts. What

C picks out– what it represents – cannot to any degree be a function of what any other concepts represent. For if what C picks out *is* a function of what other concepts pick out, then *ipso facto* C is not, contrary to our supposition, a *basic* expression in the language of thought. So if we think in a language, it follows that we have some concepts such that what any given one of those concepts picks out is to no degree a function of what any other concept picks out. This is just another way of saying that, if we think in a language, then we have some concepts that operate in complete isolation of other concepts – that pick out what they do in complete independence of other concepts. And this, in turn, means that there are foundational concepts, i.e. concepts such that possession of any one of them doesn't presuppose possession of any other concepts at all.

But there are no such concepts. Possession of any concept presupposes possession of other concepts. This is an immediate consequence of some basic facts about representation. It is impossible to represent something without representing it *as* having this or that property. It is impossible to represent something as having a property unless one can represent what is *specific* about that property. Doing this involves representing how things with that property are *different* from things without it. And doing *this*, in turn, involves being able to represent those other properties. So if one is to represent anything, one must be able to represent other things. So there are no *basic* representational capacities. For reasons just seen, if we thought in a language, there would have to be basic representational capacities. Therefore, we do not think in a language. For that matter, our thoughts cannot be mediated by anything that is even remotely language-like. For surely one of the essential properties of language is possession of the two-tiered structure just discussed; and anything lacking that structure bears no resemblance to any conceivable language. Therefore, the network of concepts that enables us

to think not only *isn't* a language, but lacks even the most tenuous resemblance to any language.

References

FODOR, J., The Language of Thought Thomas Y. Crowell 1975

FODOR, J., Concepts The M.I.T. Press 1998

HORST, S., Symbols, Computation, and Intentionality University of California Press 1996

SELLARS, W., 'Empiricism and the Philosophy of Mind' in Science, Perception, and Reality Routledge Kegan Paul 1956

WITTGENSTEIN, L., *Philosophical Remarks* edited by Rush Rhees. Translated by Raymond Hargreaves and Roger White. Basil Blackwell Oxford 1975

The Bundle Theory of Personal Identity

Abstract: According to common sense, there is more to a person's mind than the mental contents which that mind comprises: there is some 'self', underlying these contents, which *has* them (in roughly the manner in which a physical 'substance' is supposed to 'have' certain properties). Many psychologists and philosophers reject this notion, saying that, ultimately, there is nothing to a person's mind besides the mental contents which it comprises. This view is sometimes referred to as the 'bundle theory' of personal identity. A popular objection to the bundle theory is that it is circular: when asked to say under what circumstance two mental contents belong to the *same* mind, the bundle theorist (it is alleged) is forced to say that they belong to the same *self*. So the bundle theorist (it is alleged) implicitly defines a self as a bundle of mental contents that belong to the same self. Thus his analysis is circular. In this essay, I wish to show that the bundle theory is in no way guilty of circularity. For two mental contents belong to the same mind is *not* for them to inhere in some common mental substrate, but for them to have certain causal powers *vis a vis* one another. A mind can then be defined as the class of mental contents (past, present, future, possible) that bear this causal relation to a given mental content.

I. According to one point of view:

(i) A person's[245] mental contents are all there is to his mind. A person is *wholly constituted*[246] by the particular mental contents that he has. There is not some entity, some 'self' underlying any collection of mental contents.

[247] This is not to deny that there are selves. It is to say that selves are *composed* of (constituted by) the mental contents which they have.

Let us refer to this view as the 'bundle theory of personal identity.' (We are, if I am not mistaken, following customary usage here.) The alternative to the bundle theory is essentially this:

(ii) A person is *not* wholly constituted by the mental contents which he has. A person is, at least in part, something which *underlies* – or, in any case, stands apart from – his mental contents.

(i) has been criticized on a number of grounds. My objective in this paper is to show that *one* of these criticisms can be met, i.e. that one of these criticisms does not invalidate the bundle theory. It is by no means a weak criticism. (Indeed, given how much work we will have to do to neutralize it, it is clear that it is quite powerful.)

The criticism in question is essentially this:

(iii) Clearly some pairs of mental contents belong to the *same* mind while others do not. Your feeling of thirst and your perception of the ice-water belong to the same mind. By contrast, my feeling of thirst and your perception of the ice-water do not belong to the same mind. Supposing that the bundle theory is correct, what is it that makes two mental contents belong to the *same* mind? The proponent of (**) has no trouble dealing with this problem: two mental belongs to the same mind just in case they are had by the same self. But the bundle theorist cannot avail himself of this response. If he did, his position would be circular. According to the bundle theorist, a self

just *is* certain class of mental contents. So for a mental content to belong to a certain self is just for that mental content to have a certain relation to other mental contents. Let R be this relationship, whatsoever it might turn out to be. R cannot be defined in terms of the notion of f 'being had by the same self' since, for the bundle theorist, it is this very relation that R is meant to explain. So the bundle theorist must say what R is without helping himself to the notion of 'belonging to the same self.' But there is no way – or, at any rate, no non-circular way -- to do this. With regard to every attempt to define R *independently* of the notion of being had by the same self, every such attempt has failed outright or has involved some kind of covert circularity. [248]

My objective in this essay is to show that R *can* be defined independently of the notion of 'being had by the same self'. In other words, I wish to show that we can say what it is for two mental contents to belong to the same mind *without* any circularity – *without* appealing to the notion of 'being had by the same self.' In a word, we can show that the bundle theory is not guilty of circularity. [249]

By way of anticipation, my analysis is very roughly[250] as follows. It is pretty clear that two mental contents that belong to the same mind have causal powers *vis a vis* one another that mental contents belonging to different minds do not. My mental contents can interact with one another with a directness with which my mental contents cannot interact with yours. (To be sure, my mental contents *can* affect, and thus interact with, yours. But various intermediaries – e.g. sounds, overt body movements – are required to do this, and such intermediaries are *not* required for my mental contents to

affect one another.) Typically, it is thought that, if two mental contents can interact with the aforementioned directness, that is *because* they belong to the same self: in other words, their belonging to the same self is what *enables* them to interact with that kind of directness. As against this, I will argue that *what it is* for two mental contents to belong to the same self just *is* for them to be able to interact with that kind of directness. The relation of being-able-to-interact-with-a-certain-directness is primary, and the relation of belonging to the same self is derivative of thereof. A self, then, can naturally be seen as constituted by the class of all mental contents (past, present, future) that bear a certain causal relation to a given mental content.

This position may seem obvious. But it is extremely difficult to work the details associated with it. And what I do in this paper that has not (so far as I know) been done elsewhere is to work through these thorny details.

II. Suppose that I feel thirst but do not see a certain glass of ice-water, and that you see that same glass of ice-water but do not feel thirst. Ceteris paribus, the conjunction of that feeling with that perception will not result in anyone's reaching for the glass: no new mental content (no intention) will result from the co-existence of these two contents. By contrast, if *I* perceive I the glass of ice-water and *I* feel thirst, then there probably will (ceteris paribus) result a new mental entity from this co-existence of mental entities: I will intend to reach out and grab that water (or I will at least be tempted to do so).

Of course, my mental contents can affect your mental contents. If I am angry, I may yell at you, making you sad or angry. But my mental contents can affect your only through *physical intermediaries* - certain motions of my body. By contrast, it seems (at least at first glance) that for a single person's mental contents to affect one another, no such intermediaries are needed. This

suggests a certain way of defining R: ‘two mental contents belong to the same mind just in case they can directly interact.’ (We will soon see need to revise this analysis. But, I submit, it points us in the right direction.)

Before we proceed to these revisions, one point must be made clear. Given two distinct mental contents that belong to a single mind, those two mental contents may not interact; they may remain isolated from each other. Human minds are not perfectly integrated. Joe's knowledge that pi is irrational may never interact with his hatred of pistachio ice-cream; no mental content need arise from their coexistence. But that is perfectly compatible with our provisional analysis. Two mental contents, according to this analysis, belong to a single mind just in case they *could* directly interact with each other; it isn't necessary that they actually interact. Given any two contemporaneous mental contents, no matter how different, if those mental contents belong to a single mind, it will always be possible to hypothesize some circumstance under which they might directly interact to form a new mental content. (If Joe is told that either he must yell out the name of some irrational number or he will have to eat a gallon pistachio ice-cream, Joe will form an intention to yell out 'pi!' Here we actually have two cases of mental contents' directly interacting with one another. First, Joe's knowledge that pi is irrational, in conjunction with his horror of pistachio ice-cream, results in Joe's believing that he must yell out 'pi' if he is to avoid eating pistachio ice-cream. Second, this belief combines with his desire not to eat pistachio ice-cream to form an intention to yell out 'pi'.)

Unfortunately, our work has just begun. For some profound objections can be made to our analysis; and we must show that it can survive them. More accurately: we will consider some criticisms that warrant the *rejection* of our analysis. What we will now show is that an analysis very similar to that one in spirit, but slightly augmented, can survive those same criticisms.

Again, our initial analysis is this: two mental contents belong to the same mind just in case they can directly interact. But it is clear that some mental contents that cannot possibly directly interact can belong to the same mind. For example, ten years ago I had all kinds of experiences. Let A be any one of these experiences. Right now I am having various sensory experiences. Let B be any one of these experiences. A and B belong to the same mind. But clearly they couldn't possibly interact, at least not directly, since A no longer exists.

There is an obvious way to amend our hypothesis to cover such cases. Two mental contents A_1 and A_n belong to the same mind if there is some chain of mental contents A_1, A_2...A_n such that, for each i, A_i can directly interact with A_{i+1}. (If one mental entity *creates* another one, that seems to qualify as a kind of limiting case of interaction. So if my experience of eating squid leaves me with a memory of eating squid, that experience may be said to have 'interacted' with that memory by virtue of bringing it into existence.) It seems fairly clear that any two contents belonging to a single person satisfy this condition. Twenty years ago I ate squid. The experience was extremely unpleasant. I know that tonight, if I go over to Mary's house, I will be forced to eat squid (for that is what she is serving). I thus form the intention *not* to go to Mary’s house. So my experience of eating squid those many years ago has interacted with a piece of knowledge of mine to form an intention. But this interaction was not direct; for at no point did that experience coexist with that piece of knowledge. The experience left a memory of itself; and this memory interacted directly with my knowledge to form an intention. So, mediating between my experience and my knowledge, there was a chain of mental entities (in this case a chain of one - but clearly there might have been more).

The next qualification to my analysis will require quite a bit more

explaining. Our newly amended hypothesis is this: Two mental contents A_1 and A_n belong to the same mind if there is some chain of mental contents A_1, A_2...A_n such that, for each i, A_i can directly interact with A_{i+1}. We must ask: what is it for two mental entities to interact 'directly' with each other? Earlier I said that direct interaction is interaction that doesn't involve any 'physical intermediaries.' (For my mental contents to have any effect on yours, I must produce certain noises or act in certain ways: so for my mental contents to affect yours there must be 'physical intermediaries' linking them. By contrast, when two of my mental contents interact with each other, no such intermediaries seem to be involved.) But to this point, an important objection can be made:

First of all, if any form of materialism is correct, then any two mental contents are physical; and if that is the case, then whenever two mental contents interact, this interaction does *involve physical events. From this, presumably, it would follow any such interaction involves 'physical intermediaries.'*

If materialism is correct, then whenever two mental contents interact, this interaction certainly involves physical processes. But it doesn't follow these any of these processes qualify as *intermediaries.* Suppose that mental contents A and B interaction through some physical process C. If C is to be an *intermediary* process, it cannot be constitutive of either A or B; it must (so to speak) lie *between* them and must therefore not be a part of either. What I wish to show is that direct interaction between two mental contents doesn't involve any physical *intermediaries*, though (if materialism is correct) it certainly involves physical processes.

The concept of direct interaction must be analyzed one way if materialism is correct, and another way if dualism is correct. (Actually, as we will later see, these two analyses collapse into one. But it will conduce to clarity to state them separately.) In this section I will try to show (a) that if *dualism* is correct, two mental entities interact directly if their interaction involves *no* physical intermediaries. Then I will try to show (b) that if *materialism* is correct, then two mental entities directly interact if at least *some* of the physical processes constitutive of the one mental entity can affect at least *some* of the physical processes of the other mental entity without having to do so through *any* physical intermediary whatsoever - if, in other words, these two sets of processes can, to some extent at least, affect each other *not* via any intermediary processes.

A parenthetical point: (b) might seem circular: I am saying that two mental entities directly interact if the physical processes constitutive of them directly interact. So I seem to be defining 'direct interaction' in terms of itself. But what is really happening is that I am defining the obscure concept of direction interaction of *mental* entities in terms of the quite distinct, and much less problematic (in any case, much less *philosophically* problematic) notion of direct of interaction of *physical entities.* So I am not defining any concept in terms of itself.

First let us try to prove (b). Suppose I have a desire for money, and suppose that you have a strong desire to help out anyone who needs money. If I make certain noises - if, for example, I say 'I desperately need money; otherwise my life will meaningless. You're the only one who will help me' - this may well affect what mental contents you have. In particular, it may create in you an intention to give me money. But consider the physical act by means of which I got you to have this intention. This act consisted of various movements in my body which resulted in various disturbances of the

atmosphere in your vicinity. Though they are *causally* related to my desire and to your intention, these bodily movements and atmospheric disturbances have absolutely no *logical* relation to either. In other words, these physical processes do not *constitute* either of these mental contents. So, in principle, and even in fact, I could have that desire and you could have that intention quite independently of whether the aforementioned bodily and atmospheric phenomena occurred.

Now let us consider a case where two mental contents belonging to a single mind interact with each other. Suppose that my desire for ice-water interacts with my perception of a glass of ice-water to form an intention to reach out and grab the glass. If materialism is correct - as it may well be - this interaction consists entirely of some *physical* process. But *what* physical processes are involved in this interaction? Those that *constitute* the two mental contents. So two mental contents directly interact if ($\rightsquigarrow$) they interact and ($\cdot\not\cdot$) no physical processes other than those constituting those mental contents are involved in this interaction. Of course, this amounts to saying that two mental contents directly interact if they interact and nothing *other* than those mental contents are involved in that interaction - if, in effect, this interaction doesn't involve any intermediaries. (This, of course, is just what we've held all along.)

But this proposal will no doubt be met with an objection:

Suppose that a desire of mine directly interacts with a perception of mine to form an intention. Maybe the desire and the intention are 'connected' to each other via physical processes that do not to any degree constitute either of those contents - just as the flicking of a switch is connected to the subsequent illumination of a certain light-bulb via processes that are not constitutive of either of these events.

To respond adequately to this objection I must anticipate a point that I will fully develop only later on. Let A be my feeling of thirst, and let B be my perception of the glass of ice-water. Given this, imagine the following. A and B interact to form an intention to reach and grab the glass of ice-water. (Let C be this intention.) A and B are connected to each other by various physical processes. These mediating physical processes do not to any degree constitute either A or B: A and B are logically and metaphysically completely independent of these intervening physical processes.

Notice that C is not *just* the result of A and B. C is in some sense *appropriate* to A and B. If someone is thirsty and he sees a glass of ice-water, it *makes sense* for him to have an intention (or at least a desire) to reach out and grab the glass. Such an intention is normatively - if not logically or ethically - appropriate to the mental contents from which it arose. Given that C is appropriate to A and B, it seems to follow that A and B were involved in C's creation, not just in a *causal* sense, but in some more intimate sense: it seems to follow that A's *content* - *not* (merely) its content in a physical or spatio-temporal sense, but what might be called its *semantic* or *thematic* content - and B's *content* (its thematic content) are implicated in this transaction. (In the jargon of social scientists, the interaction of A and B, whereby they cause C to occur, is not merely causal but also *explicative*.)

What is true of this particular case is true of many, possibly all, cases where a single person's mental contents interact. The result of such an interaction - the new mental content that comes out of it - is *appropriate* in a thematic or semantic sense to the mental entities involved in that interaction. I see Wendy walking towards me; I dislike Wendy; I form an intention to run away. This intention is thematically or semantically *appropriate* to the mental contents whose interaction created it. It would of course be easy to multiply

examples of this principle. In fact, it would, I submit, be impossible to identify a single case where two mental contents belonging to a single person directly interacted and yet did *not* produce a new mental content that was appropriate to them.

There is nothing comparable to this in the non-mental world. Purely physical interactions[251] (i.e. physical interactions that are not also mental) do not have any normative dimension; the results of such interactions are neither appropriate nor inappropriate to the nature of the entities and processes involved therein.[252]

Let us return to our hypothetical scenario. By hypothesis, certain physical processes that do not constitute either A or B mediate between them: it is through these processes that A and B interact to produce C. Trivially, either these intervening physical processes transmit or bear A and B's respective contents or they do not. To put it more bluntly, with regard to these intervening physical processes, either they themselves have the same thematic contents as A and B - i.e. they themselves either transmit or constitute a feeling of thirst and a perception of a glass of ice-water - or they do not have such contents. Let us consider each case. Suppose that they do not have such contents, that the content of none of these processes is a feeling of thirst or a perception of ice-water. Remember these intervening processes result in an intention to reach out and grab the ice-water; remember also that this intention is, in some very profound way, *appropriate* to the feeling of thirst and the perception of the ice-water. If the aforementioned intervening processes did not bear a feeling of thirst or a perception of ice-water, then it would be an absolute miracle that they resulted in a mental content that was *appropriate* to such a perception and desire. In other words, if those intervening processes did not *transmit* the *content* of the feeling of thirst or

the *content* of the perception of the ice-water, then it would be completely inexplicable why those intervening processes resulted in a new content (namely, an intention to drink the ice-water) that was so obviously appropriate to the aforementioned contents.

An analogy might help. Suppose you and I are talking on the phone. Obviously what comes out of the receiving at your end is very similar to what goes into the mouth-piece at my end. We explain this similarity by citing the fact that there is a telephone wire transmitting an electrical impulse initiated by the sound of my voice that is retrieved and re-converted into sound in your ear-piece. Obviously if the telephone wire (or the satellite hook-up) were severed, then, although it would not strictly speaking be *impossible* that what you hear at your end resembled what I say at my end, it would nonetheless be extraordinarily improbable that this should happen – if it were to happen, it would be a virtual miracle. An exact analogue of this holds in the case where my thirst and my perception of the ice-water interact to form an intention to drink the ice-water. As a result of my having the perception of the ice-water and the desire to drink it, a new mental content that is *appropriate* to these two contents is formed, viz. an intention to drink the water. The feeling of thirst and the perception are what *cause* the intention to arise (in any case, they obviously bear *some* of the causal responsibility for this). Now suppose (what may in fact be true) that materialism is correct, and that consequently the thirst and the perception exert this causal influence through *physical* intermediates. Trivially these intermediaries either do or do not transmit the contents of the feeling of thirst and of the perception. Let us consider each case. If they do *not* bear these contents, then it becomes inexplicable why it is that these processes should result in a new mental entity (viz. an intention to drink the ice-water) whose content is *appropriate* to that of the feeling of thirst and the perception of the ice-water – just as, if the telephone line (or

satellite hook-up or whatnot) mediating between my telephone and yours is severed (destroyed), then it becomes inexplicable why the sounds going into my mouthpiece should resemble those leaving your ear-piece. So we must assume that these intervening processes do bear these contents. Now if they *do* transmit this content, then *ipso facto* they *bear* it; and if they bear it then *ipso facto* these intervening processes – while physical – are *mental.* (For as we noted, if something non-derivatively bears content it is *ipso facto* mental, whatever else it is.) In which case, we have two mental entities *directly* interacting. So once again, the feeling of thirst and the perception of the ice-water satisfy our definition of belonging to the same mind: for although they themselves do not directly interact, there is nonetheless some chain of mental entities A_1, $A_2 \ldots A_n$ such that the feeling of thirst and the perception are in direct causal contact with the terminal links in this sequence and such that, for each i, A_i directly interacts with A_{i+1}. And, as we saw a moment ago, if we make that assumption, then our criterion of belonging to the same mind is satisfied.

I admit that the foregoing was dense and therefore to be regarded with suspicion. Perhaps the following will render it more perspicuous and less suspicious. In general, if mental contents interacted with each other through the medium of physical processes that did not themselves bear these contents, then it would be absolutely inexplicable how it is that, so often, when mental contents interact, the results of this interaction are thematically *appropriate* to the thematic contents of the participants in this interaction. So it cannot reasonably be said that the physical processes mediating between A and B do not themselves bear A and B's contents. Now trivially, either these intervening processes are themselves constitutive of A and B or they are not. (For x to be 'constitutive of' y is, roughly, for x to be a part of y, or for x to make up y.) Let us consider each case. If these intervening processes are

constitutive of A and B, then we have proved our point, namely, that A and B interact directly (not through any processes *other* than such as constitute them).

But if these intervening processes *are not* constitutive of either A or B, it by no means falsifies our analysis. Remember, our analysis is this: A and B belong to the same mind if they are both links in some chain of mental states such that, given any two adjacent links in this chain, those two links interact with each other directly, i.e. not through the medium of any intermediaries whatever, be they physical or mental. At least some of the physical processes bearing A's content must directly interact with some of the processes bearing B's content. We have already seen why this must be so: if it doesn't so, then it becomes impossible to explain why a mental content that is thematically or semantically *appropriate* to A and B arises from their coexistence. So even if A and B do not themselves directly interact, we must assume that some processes transmitting A's content directly interact with some processes transmitting B's content. Of course, these intervening processes are mental, even if they are also physical. (I will justify this claim in a moment.) Moreover, we must assume that some of the processes transmitting A's content directly interact with A and that some of the processes transmitting B's content directly interact with B. (Again, if we do not assume this, then we cannot explain why A and B share their contents with these intervening processes.) So A and B satisfy our criterion for belonging to the same mind. A directly interacts with some mental processes that bear A's content; these latter contents directly with some that bear B's content; and, finally, these latter directly interact with (are produced by) B. So A and B are terminal links in some chain of mental states or events such that any two adjacent links in this chain directly interact with each other.

I said a moment ago that at least some of the processes mediating between

A and B must themselves be mental. What is the justification for this claim? These intervening processes transmit thematic or representational content. (They transmit A's content and B's content.) Moreover, their power to do so doesn't derive from human convention or artifice: so in this respect they are unlike paintings or sentences, whose intentionality is derivative. Surely anything that non-derivatively bears thematic or semantic content is mental: surely that is sufficient, even if it is not necessary, for being mental. So these intervening processes are mental.

To sum up: we have shown that, even if the interlocutor is right, our analysis still goes through. Even if no process constitutive of A directly interacts with any process constitutive of B, and even if the processes mediating between A and B are physical, we are still forced to hold that *some* mental processes directly interact with some other mental process: for we are forced to hold that the physical processes transmitting A's content interact without any physical intermediaries with those physical processes transmitting B's content. We are forced to hold this because, if we do not hold it, we cannot explain why a mental content that is *appropriate to* A and B arises from their coextensive.

So our analysis still stands: two mental contents, A_1 and A_n, belong to the same mind if they are terminal links in some chain of mental contents $A_1...A_n$ such that, for each i, A_i and A_{i+1} directly interact; and (assuming materialism is correct) two mental contents directly interact if at least some of the physical processes constitutive of the one directly interact with those constitutive of the other; in other words, if, at least to some extent , these two mental contents interact *not* through any intermediaries. (Remember that, although it *looks* as though I am defining 'direct interaction' in terms of itself and that my analysis is therefore circular, I am not defining anything in terms of itself. I am defining the problematic concept of direct interaction for

mental states in terms of the less problematic - and, in any case, different - concept of direct interaction for *physical* states.) So two mental contents directly interact with each just in case they can (at least to some extent) they affect each other *not* via any physical intermediaries (i.e. not via any physical processes that lie *between* them but are constitutive of neither of them).

This definition rules out exactly the kinds of interactions that we want to rule out and rules *in* exactly the kinds of interactions that we want to rule in. My mental contents can affect your mental contents. But none of the processes constitutive of any of my contents interacts directly with any of the processes constitutive of any of your mental contents. So, our analysis ascribes my mental contents to one mind and it ascribes yours to a separate: and this is just what it should do. On the other hand, given any two of my mental contents, it does seem possible *either* (i) that their interaction should involve the *direct* interaction of some processes constitutive of the one content with some of those constitutive of the other mental content; or (ii) that it should be possible to interpose some chain of mental contents between them such that each link in this chain directly interacts with its neighbours. So our analysis ascribes these contents to the same mind: and this, of course, is as it should be.

So far we have said what it is for two mental contents to interact directly *assuming* that materialism is correct. Now let us defend analysis (a). In other words, let us try to prove that, if *dualism* is correct, then two mental contents directly interact if their interaction involves no physical processes - and a fortiori no physical intermediaries - whatsoever.

Suppose ($\sim$) that dualism is correct and ($\cdot\not\cdot$) that for two mental contents to interact with each other, physical intermediaries *are* necessary. Given that dualism is correct, it follows trivially that no physical event is identical with, or constitutes, any mental event. So, assuming ($\sim$) and ($\cdot\not\cdot$), it

follows that something *non*-mental must mediate between any two mental contents.

At first it might seem that this non-mental something could be a physical process that *represented*, but wasn't identical with or constitutive of a mental content. (Of course, this physical process would have to be *non-derivatively* intentional. If there are brain processes mediating between mental states, surely they are intrinsically intentional, and not intentional in the second-class sense in which sound tokens and ink-deposits are intentional.) But, in fact, this would not be compatible with dualism. As we noted earlier, if a physical process non-derivatively represents a mental content, then *ipso facto* that process is, at least to that extent, itself mental. So if (≁) and (·⌿·) are correct, then mental contents can interact with each other only through physical processes that do not intrinsically represent mental contents. But, as we have now seen many times over, unless we assume that mental contents interact with each other through processes that themselves have mental content, it would be absolutely impossible to explain why it is that, so often, when mental entities interact, the result is thematically or semantically *appropriate* to the participants in that interaction; in other words, unless we make the aforementioned assumption, we cannot account for the normative dimension of mental interaction. So if we assume (≁) that dualism is correct, then we cannot assume (·⌿·) that mental entities interact with each other through the medium of physical processes.

Let recapitulate what we have established. If dualism is correct, then physical processes are not intrinsically representational. If they aren't intrinsically representational, then we cannot assume that interaction among mental states is always mediated by physical processes; for if we did make this assumption, then we would have no way of accounting for the normative dimension of certain cases of mental interaction. So if dualism is correct, then

two mental contents *directly* interact if their interaction involves *no physical intermediaries whatsoever.*

Of course, if A is any mental content belonging to me and B is any mental content belonging to you, then any interaction of these two contents *will* necessarily involve physical intermediaries. So our analysis correctly ascribes A and B two different minds.

What about if A and B belong to the same mind? Will our analysis ascribe them to the same mind? We have seen that if two mental contents belong to the same mind, then it is possible for them to interact in a normatively or semantically significant way. We have also seen that, whenever two mental contents interact in this way, it is because any events mediating between them bear their respective contents. If dualism is correct, these mediating events must be mental (since, as we've seen, if dualism is correct, physical processes cannot non-derivatively transmit mental content). So we must assume that, given any two mental contents that interact in a normatively significant way, these two contents are mediated, if at all, only by other mental contents - in other words, they are mediated, if at all, by a chain of mental contents such that each link in this chain interacts *directly* (through no intermediaries, mental or physical) with its neighbours. This concords with our analysis. For that analysis, it will be remembered, is this: A and B belong to the same mind just in case they are link in a chain such that each link in that chain interacts directly with its neighbours. So our analysis ascribes two mental contents to the same mind when, in fact, they belong to the same mind.

We have given two separate analyses of what it is for two mental contents to interact directly. But it is clear that these two analyses really collapse into one. If materialism is correct, two mental contents directly interact just in case the physical processes constitutive of the one affect those constitutive of the other without there having to be any *intervening* physical events.

(Physical events are involved, but only such events as *constitute* the two mental contents in question; so no physical events *intervening* between the two contents are involved.) So, if materialism is correct, two mental contents directly interact just in case there are no *intervening* physical events. If dualism is correct, then two mental contents interact directly just in case *no* physical events - and a fortiori no intervening physical events - are involved in their interaction. So, regardless of whether materialism is correct or not, two mental contents directly interact just in case their interaction doesn't involve any *intervening* physical events.

Now we are in a position to give a bundle-theoretic definition of what a self is. Two mental contents, A_1 and A_n, belong to the same person just in case there is some chain of mental contents A_1, A_2...A_n such that, for each i, A_i interacts directly with A_i+1. (If one mental contents creates another, that may be regarded as a limiting case of interaction.) Let us say that A_1 and A_n are *I-related* to each other. A self or person may be defined as the class of all mental contents I-related to a given mental content. [253]

Chomsky, Marr, and other cognitive scientists have given us reason to believe that there are deeply unconscious mental processes: mental processes that (unlike those belonging to the Freudian unconscious) cannot possibly be made conscious. (For example, the computational processes mediating language-learning and visual processing simply couldn't be made conscious. They are part of a framework that underlies consciousness – that supports it – and that cannot, therefore, itself be conscious.) Let P_1 be one of these deeply unconscious processes that is occurring in my mind, and let P_n be some conscious content of mine. Obviously (by assumption, really) P_1 and P_n belong to the same mind. But I don't see how your analysis can accommodate

this fact. For I don't see that there could be any chain of mental contents P_1, $P_2 \ldots P_n$ *such that, for each i,* P_i *directly interacts with* P_{i+1}.

(Let us suppose for the sake of argument that there are such 'deeply unconscious' process: for obviously if there are not, then the objector's point becomes null and void) The objector's point is singularly misguided. With regard to the deeply unconscious processes in question, though they cannot themselves be conscious, they obviously have *effects* on our conscious contents. Indeed, items of consciousness are supposed, according to these models, to be the results of deeply unconscious processes: the latter are posited (among other reasons) precisely in order to explain why we have the items of consciousness that we do (why have we certain conscious visual perceptions, why we have certain conscious linguistic intuitions, why we experience certain chords as dissonant and others as consonant, and so on). Further, the connections obtaining between items of consciousness and the deeply unconscious processes which, according to the view in question, create them is supposed to be *explicative* and not (merely) causal; in other words, the contents of the items of consciousness are supposed to be *appropriate* to the contents of the said deeply unconscious processes. So, if the view of Chomsky, Marr, and cognitive scientists generally, are to be correct, then there must be *content-preserving* connections holding between deeply unconscious mental entities and items of consciousness. (More specifically, these deeply unconscious processes are supposed to be *computational*; they are supposed, in effect, to embody the making of certain kinds of *inferences* – a computation is a special kind of inference: a purely formal (syntax-driven) inference. So if the models in question are correct, then our conscious mental items are the conclusions of sequences of inferences beginning with the aforementioned deeply unconscious processes.

From this it follows trivially that our conscious mental contents are in some sense -- actually, in this case a very strict sense: an inferential or logical sense -- *appropriate* to our deeply unconscious processes.) The transmission of content, as we've seen, involves *direct* interaction between mental entities. (For as we've noted, if an entity is content-bearing it is ipso facto mental; and, as we've also seen, if a non-content bearing entity mediates between two content-bearing ones, then if the content borne by the first of those entities *is* picked up by the second, it can only be through some accident or coincidence.) So if the psychological models in question are correct, then – in direct contradiction to what the objector says -- it is actually *de rigueur* to suppose that, given any deeply unconscious content C_1 and any conscious content C_n, such that both belong to the same mind, there is some *possible* chain (if not an actual one) of mental contents C_2, $C_3 \ldots C_{n-i}$ such that C_1 directly interact with C_2, C_{n-1} directly interacts with C_n, and for each i, C_i directly interact with C_{i+1}.

We have, I think, shown a way to meet *one* criticism of the bundle theory. But there are others. And we have provided no *positive* argument for that theory. Right now I would like to briefly show that one of these other criticisms is misguided; and, in doing this, we will find *positive* reason to think not only that the bundle theory is reasonable, but that it is all but analytically true.

The criticism is this:

The bundle theory says that a person's self is nothing but the mental contents which that person has. Now that is the same as saying the following. Let C be the set or class of some person's mental contents. That person *is* C.

There is an obvious problem with this. Like all sets, C is "modally frozen." Let C* be some set that is just like C except that C* contains one

mental entity that C does not have. C* is ipso facto a *different* set from C. Basically, sets are individuated by their memberships. So a set *couldn't* have had members other than those it *actually* has.

But a person obviously *could have* had mental contents different from those she actually has. In the metaphysical (though perhaps not the deterministic sense of "could have") I could have watched "Love Boat" ten minutes ago, instead of reading Recanati's book *Literal Meaning*. In that case, I would have had different mental contents from those I in fact had.

So people are not modally frozen; anyone *could* have had different mental contents from those she actually had. But sets are modally frozen; a set *could not* have members other than those it in fact has. So people are not sets of mental contents. So people are not bundles of mental contents. So there is something to a person *over and above* the mental contents which that person has – some self which *underlies* them and is distinct from them.

The logic here is technically good, but superficial; and this argument demands a mere change of nomenclature. Consider the following analogue of that argument.

There is something to the nation of France *over and above* the thoughts and deeds of the various people who compose that country. So France is something quite distinct from the set of those thoughts and deeds. Why? Let C be the set of all those thoughts and deeds. And let C* be some set very similar to C, except that C* contains one thought or deed not contained in C. In that case, C and C* are different sets. But surely some French person *might* have thought or done something other than what he actually did. So France is not modally frozen. Thus France is not a set of any kind. Thus France is not the set of deeds and thoughts had by French people. Thus there

is something to France over and above the activities of the people making up that nation.

For the reason just stated, France cannot be *identified* with the set containing all and only the thoughts and deeds of citizens of France. But surely France is not something over and above those thoughts and deeds. For it is impossible, in the strictest sense of the word, that France in world w should differ from France in w* unless some *individual* in France-in-w differs from his counterpart in France-in-w* (or fails to have any counterpart at all in France-in-w*). Obviously, if the individuals composing France-in-w are just like the individuals composing France-in-w*, then those two Frances are qualitatively identical. At the same time, for the reason the objector states, France is not *identical* with the set of deeds and thoughts associated with French individuals.

The relation is one of *supervenience*, not identity. Once it is fixed what each last French person has felt, thought, and done, it is fixed what properties France itself has. Facts about France supervene on facts about the biographies of French individuals. So the bundle-theory is true of the nation of France: there is nothing to France over and above facts about French individuals, and their interactions – facts about France are determined, in the strictest sense of the word, by facts about individuals.

The bundle theory of personal (as opposed to national) identity is true in exactly the same way. For the reason the objector states, a person is not *identical* with any set. But if two people P and P* are identical in respect of what their mental contents are, then it is impossible, in the strictest sense of the word, that there should be *any* qualitative differences between them. Once the facts about a person's mental contents are settled, *everything* about that

person is ipso facto settled. So there is no reason to posit a *self* distinct from those mental contents. To do so would be to posit idle machinery; it would be like a positing a self *underlying* the activities of various French individuals – that self being the *true* nation of France. But such a manoeuvre would be absurd – would be metaphysical extravagance. There are various French people, doing and thinking various things. There is not some ur-entity – some Hegelian uber-geist – of which these various activities are manifestations. And positing some such ur-entity would be explanatorily worthless: it wouldn't give us the "cohesion relation" that makes two acts both be constitutive of the nation of France. Surely we must look for that cohesion relation in some *causal* connection between the two acts – the cohesion relation holds between the acts *themselves,* and not by way of some mysterious national "self" underlying those acts. (Of course, other *acts* or states of affairs might be involved in that cohesion relation; but not, surely, some ur-entity – some national under-soul or whatnot.) The same logic applies to individual selves. It would be non-sensical to say: P and P* are qualitatively different people, but they are qualitatively identical in respect of their mental contents. So it would be idle, if not absurd, to posit a self underlying mental contents. We *can* (I think we've seen) explain the cohesion relation without positing some underlying self. And we *must.* Given that people cannot *differ* (qualitatively) unless they differ in respect of some specific mental content, it follows that everything constitutive of a person lies in her mental contents and, thus, that nothing constitutive of her lies in anything *other* than those same contents: so there is no "self" distinct from, or underlying, a person's mental contents.

References

Armstrong, David

[1] A Materialist Theory of the Mind Routledge and Kegan Paul 1968. P. 17.

Blackburn, Simon

[1] Ruling Passions

Carruthers, Peter

[1] Introducing Persons Croom and Helm 1986

Hume, David

[1] A *Treatise Concerning Human Nature* Edited, with an Analytical Index, by L.A. Selby-Bigge; Second Edition, with text revised and variant readings by P.H. Nidditch. Oxford at the Clarendon Press 1978

Olson, Eric
[1] The Human Animal, Oxford University Press 1997

Parfit, Derek

[1] Reasons and Persons Oxford University Press 1984

Sartre, Jean-Paul

[1] *The Transcendence of the Ego* : an existentialist theory of consciousness / Translated and annotated with an introduction by Forrest Williams and Robert Kirkpatrick. -- New York : Noonday Press, [1957].

Stern, Cindy

[1] 'Hume and the Self at a Moment' in History of Philosophy Quarterly volume 4, number 2, April 1987

Williams, Bernard
[1] ‘The Self and the Future’ In *Personal Identity*, ed. John Perry. Berkeley, CA: University of California Press, 1975, pp. 179-98.

Functionalism incompatible with the causal efficacy of the mental *qua* mental

Abstract: According to functionalism, for x to be an instance of a particular mental type (e.g. the type *belief that snow is white*) just *is* for x to have certain causes and/or certain effects. Functionalism thus makes it impossible to explain how the representational and phenomenal contents of mental entities *ground* their causal properties. A consequence of this is that, if functionalism is right, no sequence of mental states constitutes an *inference*

of any kind: functionalism allows that there are sequences of states that *conform* to logical rules, but does not allow (so I will argue) that those sequences embody a *grasp* of those rules.

Given any type of mental entity, instances of that type characteristically have certain causes and effects. A belief that $2^2=4$ is likely to result from certain things (e.g. certain sense-perceptions conjoined with certain beliefs and certain cognitive faculties) and is likely to result *in* certain things (e.g. the belief that 2 has an even square).

A doctrine known as *functionalism* says: for something to be the belief that $2^2=4$ just *is* for it to have one of those causes or effects. In general, according to functionalism, for x to be instance of mental type T just *is* for x to have certain causes or certain effects.

Another illustration may be in order. A certain kind of sense-perception (i.e. sense-perception with a certain phenomenology and representational content) will characteristically have certain causes or certain effects. According to functionalism, for something to be that sort of sense-perception just *is* for it to have one of those causes or one of those effects. Put another way, x is (e.g.) a perception of such and such a type *in virtue of* having certain causes/effects.

There are excellent reasons to be a functionalist. These are sufficiently well known that they needn't be stated here.[254]

The purpose of this note is to expose what I think is a serious defect on the part of functionalism – a defect that, it seems to me, warrants the complete rejection of that doctrine.

Suppose that you are driving along and you have a sense-perception of a

stop-sign. Let us assume that you are cognitively normal, that you have assimilated the mores characteristic of our culture, that you do not wish to be stopped by a police-officer, that you know this particular intersection to be well patrolled by police-officers, and so forth. Under these circumstances, the perception is overwhelmingly likely to induce you to stop driving. More exactly, it is likely to result in an intention to stop driving (various things – e.g. nerve damage – may interfere with your discharging this intention). Surely this is just the kind of cause-effect relationship that functionalists have in mind when they say that for x to be (e.g.) a perception of such and such a type just *is* for x to have certain causes or effects.

The question arises: *why* does the perception induce you to stop? It causes you to stop because it has a certain content: it bears certain information (roughly, it tells you that there is a certain kind of object in front of you); and on the basis of this information, you think there may be unpleasant consequences if you don't stop. So the perception causes you to stop *because* it is a perception of such and such a kind – *because* it is a perception with such and such representational content.

Of course, all of this is obvious, even trivial. Representational mental entities have the effects they do *because* (at least in part) they have a certain representational content. Your perception of the rabid dog causes you to flee *because* that perception has a certain content: it tells you that there is a rabid dog present; and you know that, because of that fact, you should flee. This point applies *mutatis mutandis* to any mental entity that has representational content. And I submit that a serious problem for functionalism lies in this innocuous point.

Before proceeding we must be careful about one thing. The word 'because' has two very different meanings. It has what might be called a 'causal' meaning and also what might be called a 'logical' meaning. In the

sentence

(*) ‘there is smoke because there was a fire’,

the word ‘because’ has a causal meaning; that sentence more or less means ‘the cause of the smoke was a fire’.

In the sentence

(**) ‘x has interior angles adding up to 180° because x is a triangle’,

the word ‘because’ does *not* have a causal meaning. It is not as though x was a triangle at time t1 and that *caused* it to have interior angles adding up to 180° at time t2. Such a claim would be nonsensical. Instead of saying (**), one ought to say

(***) ‘x has a interior angles adding up to 180° *in virtue of* the fact that it is a triangle’

The ‘in virtue of’ locution eliminates the misleading causal connotations borne by (**).

We noted a moment ago that one stops at the stop-sign *because* one has a perception of such and such a kind of (i.e. possessed of such and such a representational content). How is the word ‘because’ being used here? Surely it is being used in the *causal* sense. Your having a perception with a certain representational content is what *causes* you to stop.

I wish to be as clear on this point as possible (it will soon become clear why). Let P1 be your perception of the stop-sign. The *cause* of your stopping is *P1’s having a certain representational content.* So what causes you to stop

is *P1's having a certain representational content.*

I do not see how this could be contested. Nor do I see how it could fail to generalize to virtually every other representational mental entity. Let P2 be your perception of the rabid dog. Surely the *cause* of your fleeing is (inter alia) P2's having a certain representational content (in effect, it is telling you that there is a certain kind of dog in front of you). Let P3 be your perception of a tall glass of ice-water in front of you (let us assume you are thirsty, know the water not to be poisoned, and so on). Surely the *cause* of your drinking the water is (inter alia) P3's having a certain representational content.

But what does *functionalism* say about the relationship between your seeing the stop sign and your subsequently stopping (remember that P1 is your perception of the stop-sign)? We know what it *should* tell us: functionalism *should* tell us that it is P1's having such and such representational content that causes you to stop. But what does functionalism actually tell us?

Functionalism tells us that us that P1 has the representational content that it has *in virtue of* (inter alia) causing you to stop. So P1 is a perception of (e.g.) a sign of a certain shape and color, a certain distance from one, under such and such lighting conditions, and so forth, *in virtue of* (inter alia) its causing one to stop. (To simplify discussion, let R1 be the representational content of P1 – so R1 is, we may suppose[255], information to the effect that an object with a certain shape and color is a certain distance from one, and so forth.) So functionalism says that P1 has R1 *in virtue of* the fact that (inter alia) P1 causes you to stop.

Now if P1 has R1 *in virtue of* the fact that P1 causes you to stop, then P1's *having* R1 cannot be what *causes* you to stop. If x is to be a cause of y, then x must be distinct from y, and x must exist prior to (though perhaps contiguously with) y. We agreed that P1's having R1 is genuine *cause* of

your stopping. Functionalism says: P1 has R1 *in virtue of* causing you to stop; P1's having R1 is a kind of consequence (not a causal consequence, of course) of P1's causing you to stop. So, if functionalism is right, P1's having R1 cannot be a *cause* of your stopping.

To elucidate this: if x holds *in virtue of* y, then x cannot *cause* y. If x is to cause y, then x's existence must be entirely independent of y's: otherwise x will not be able to *precede* y, and therefore won't be able to cause y. So if P1 has R1 *in virtue of* causing you to stop, then P1's having R1 cannot be a *cause* of your stopping. If causing you to stop is what makes P1 have R1, then plainly P1's having R1 cannot be what causes you to stop. If y is responsible for x's being the case, then x's being the case cannot cause y to happen. Analogously, if P1's causing you to stop is responsible for P1's having R1, then P1's having R1 x cannot *cause* you to stop.

But it is as plain as day that P1's having R1 *is* what causes you to stop. You stop *because* you see a stop-sign. Functionalism says, in effect, you see a stop-sign because you stop.

Functionalism thus gets the order of explanation the wrong way about. And functionalism gets the order of explanation wrong not just in the single case we've discussed, but in *every* instance of 'intentional' causation one might think of.

Let P2 be your perception of the rabid dog. Let R2 be the representational content of P2 – so R2 is, we may suppose, information to the effect that there is a dog that is foaming at the mouth that is such and such distance away from you, and so forth.

Your having P2 causes you to flee. And clearly your fleeing is *caused by* (inter alia) P2's having R2: what *causes* you to flee is obviously that your perception (P2) is a perception of a growling, vicious dog.

Functionalism says that P2 has R2 *in virtue of* causing you to flee. For

functionalism, P2 has R2 *because* P2 causes you to flee. In other words, P2 has R2 *in virtue of* the fact that P2 causes you to flee. As we noted, if x exists *in virtue of* y's existing, then x cannot *cause* y to exist. So, for functionalism, P2's having R2 cannot be a *cause* of your fleeing.

Functionalism says that P2's causing you to flee *grounds* or *underlies* P2's having R2. Now things cannot cause the things that underlie them. So, if functionalism is right, the P2's having R2 cannot be a cause of your fleeing. But obviously it is (or could be) a cause of your fleeing.

To sum up, you flee *because* you have a perception with a certain content -- your seeing a growling dog *causes* you to flee. Functionalism says, in effect, you see a growling dog *because* you flee. So functionalism gets the order of explanation wrong.

Before moving on, let us apply these points to the cases of itches and pains. These days not too many people are functionalists about itches, pains, tickles, and so forth. ('Phenomenal' content is widely held to resist functionalist analysis, even by those who are otherwise hardline functionalists.[256]) Nonetheless, if we apply what we said above to these things, it will render more perspicuous the structure of the argument given above.

A sensation is, pretty much by definition, something which has a phenomenology. Indeed, for something to be a sensation essentially *is* for it to have a certain phenomenology. And for something to be a pain or a tickle or an itch (of a particular type) just *is* for that thing to have a particular phenomenology.

The functionalist says that x's causing you to scratch *is* (inter alia) what makes x be an itch – so x is an itch *in virtue of* (inter alia) causing you to scratch. Now, as we just noted, for something to be an itch just *is* for it to have a certain phenomenology. So, if the functionalist is right, x has a

particular phenomenology *in virtue of* its (inter alia) causing to you scratch. If this is right, then x's having that phenomenology cannot be what *causes* you to scratch. Remember what we said earlier: if S2 exists *in virtue of* S2, then S1 cannot *cause* S2 – not even a partial cause. If x's being an itch exists *in virtue of* x's leading you to scratch, then x's being an itch cannot be a *cause* of your scratching.

But it is as clear as day that it is, not merely the itch, but the itch's having a certain phenomenology that is a *cause* of your scratching. And if the x's having a certain phenomenology is what causes you to itch, then it cannot be in virtue of x's causing you to itch that x has a certain phenomenology. In other words, the functionalist analysis of what it is to have a certain phenomenology – of what it is to be an itch – fails. Obviously, exactly an exactly similar criticism applies to the

Before moving on, I should head off one possible confusion:

Let x be some mental entity – say, a belief that arsenic poisons, and let Joe be the person who has x. And let us suppose that x causes Joe to do such and such (e.g. to throw out a jug of water he believes to be arsenic-laced). You have made a big deal about the difference between *x* being the cause of Joe's throwing out the water, on the one hand, and *x's being a belief that arsenic poisons* being the cause of this. But this is a purely verbal distinction. x is, after all, a belief that arsenic poisons; so I don't really see what difference, apart from a purely nominal one, there is between x's doing such and such, on the one hand, and x's *being a belief that arsenic poisons* doing such and such, on the other hand.

The difference mentioned here is not purely verbal. Any given entity or state of affairs will typically have a lot of different properties, and not all of these

properties are implicated in a given causal transaction involving that object. The cause of the rock's breaking the window is (inter alia) the rock's having a certain momentum, not the rock's having certain thermal properties or a certain color. On the other hand, the cause of the rock's having an algae growth on it is (inter alia) the rock's having certain thermal properties, not its having a certain momentum. Supposing that x is a belief that arsenic poisons, and that x causes such and such, it doesn't follow that x's *being* a belief that arsenic poisons is what causes such and such. Suppose that x, being a brain-state, has certain thermal properties and thus raises the temperature inside my cranium by some small amount. In that case, *x* is causing the temperature in my cranium, but x's being a belief that arsenic poisons is not doing that. So there very much *is* a distinction between x's causing me to do such and such, on the one hand, and x's *being a belief that arsenic poisons* causing me to such and such, on the other. x, like any other spatio-temporal entity, has a multiplicity of properties; among these is the property of being a belief that arsenic poisons. But it has various others: being a brain-state, it has various chemical and thermal properties, for example. And, strictly speaking, it is meaningless to say simply '*x* caused such and such' – we must cite the specific fact about x, the specific property had by x, which caused such and such. (No self-respecting physicist would say simply 'the moon is responsible for the tides'; he would say 'the moon's having such and such mass, and being thus and distance, relative to an object with such and such properties… causes such and such.') So, yes, it is true that my discussion leans heavily on the distinction between 'x caused such and such' and 'x's being a belief that arsenic poisons caused Joe to throw out the water.' But this is a distinction that bears the weight I'm putting on it.

Let us move on. Oftentimes mental entities *justify* the things which they cause – this is especially apparent where the making of inferences is

concerned – and functionalism breaks down in a fairly clear way in connection with this sort of phenomenon.

Suppose you believe that

(P) snow is white

and you also believe that

(Q) white things reflect light.

Under these circumstances, no doubt, you will have some tendency to infer

(R) that snow reflects light.

So your believing P and Q will no doubt cause, or tend to cause, you to believe R. Moreover, your believing P and Q will *justify* your believing R. So in this case – and this is by no means exceptional – mental causation coincides with justification.

We must be careful about one thing. For an inference to be made is not just for one belief to cause another belief to occur: the *contents* of those beliefs must be implicated in a certain way. My believing P and Q could *cause* me to believe R without my *inferring* R from P and Q. My brain might just be wired in such a way that, if I believe P and Q, then I automatically believe R. (Or a belief in P and Q might generate a belief in R by way of purely associative or emotional, as opposed to rational, processes.) For me to *infer* R from [P and Q], I must see the *bearing* of [P and Q] on R. So for me to really *infer* R from [P and Q], a grasp of the content of [P and Q] must be implicated in a certain way in the generation of the belief R.

So suppose that B1 is my belief in the conjunctive proposition [P and Q], and that B2 is my belief in R. For me to really infer R from [P and Q], it is not enough that B1 cause B2 to occur. It is necessary, also, that B1's *being a belief that [P and Q]* be causative of B2. (In other words, B1's being a belief that snow is white and white things reflect light must be causative of the belief that snow reflects light.) For if B1's being that kind of belief isn't operative in the generation of B2, then the representational content of B1 – viz. the conjunction [P and Q] – won't be implicated in the kind of way that it must be if the generation of B2 is to qualify as an *inference* from B1, rather than a mere *product* of B1. To put this another way: if the transition from B1 to B2 is to constitute an *inference* from [P and Q] to R, then surely it is necessary, not merely that *B1* cause B2, but that *B2's having a certain representational content* cause B2: to wit, it is necessary that B1's *being a belief that [P and Q]* be operative in producing B2. If B1'a being a belief that P and Q is not causative of the occurrence of B2, then B2 obviously didn't constitute an inference of any kind from the content of B1.

What does functionalism say about the kind of situation where one belief not only causes, but also *justifies*, the occurrence of some other belief? For functionalism, x qualifies as the belief that (e.g.) snow is white *because* (inter alia) of its effects. x qualifies as the belief that snow is white *because*, when conjoined with certain other beliefs (e.g. the belief that white things reflect light), there are certain effects (e.g. the generation of a belief that snow reflects light). So for functionalism, what *makes* B1 be a belief that [P and Q] is (at least in part) that it, B1, *causes* the generation of a belief that R. So, for functionalism, what *makes* B1 be a belief to the effect that snow is white and white things reflect light *is* (inter alia) the fact that B1 generates a belief that snow reflects light.

But if what *makes* B1 be a belief that [P and Q] is that it, B1, *causes* the

occurrence of a belief that R, then one cannot really infer R from P and Q. As we said a moment ago, for the transition from B1 to B2 to constitute an *inference*, that transition must embody some understanding of the bearing of [P and Q] on R. So for the transition from B1 to B2 to constitute an inference from [P and Q] to R, it is not enough to that *B1* cause B2: it is necessary that B1's *being a belief that [P and Q]* be causative of B2. But, for functionalism, B1's causing B2 *underlies* B1's being a belief that P and Q. In other words, for functionalism, B1's causing a belief that snow reflects light *underlies* B1's being a belief that [snow is white and white things reflect light]. Put yet another way, for functionalism, B1 is a belief that P an Q *in virtue of* (inter alia) causing B2.

But recall what we said earlier. A thing cannot *cause* anything that underlies it; x cannot *cause* y if y underlies x or (what is the same) if x exists *in virtue of* y. For functionalism B1 is a belief that [P and Q] exists *in virtue of* the fact that B1 causes a belief that R. It follows that, if functionalism is right, then B1's *being a belief that [P and Q]* cannot be a cause of the belief that R. But if B1's being a belief that [P and Q] is not operative in bringing about the belief that R, then the latter belief was not arrived by way of *inference* from the former. As we've said, for the transition from B1 to B2 (the belief that R) to be an *inference*, rather than a case of garden-variety causation, B1's *being a belief that P and Q* must be causally efficacious in that transition. But if functionalism is right, it is in virtue of B1's leading to B2 that B1 is a belief that [P and Q]. Therefore, if functionalism is right, B1's being a belief that [P and Q] cannot be a cause of B2 (if x exists in virtue of y, x cannot cause y to exist). So functionalism doesn't allow the generation of one belief out of another to embody an act of *inference*: functionalism cannot accommodate the existence of inferences, and is therefore false.

Writer-based vs. Reader-based Prose: Analytic Summaries of three articles thereupon

Analytic Summary of Linda Flower's "Writer-Based Prose: A Cognitive Basis for Problems in Writing."

This article focuses on the distinction between "writer-based" and "reader-based" writing. Some background points may help us define these terms.

Initially, when you learn about some new domain, e.g. organic chemistry or Bolivian military history, you aren't able to see the logical and causal relations that hold among the data that you are assimilating. One always begins by learning "brute facts." It is only later, after one has acquired a certain mastery of the domain in question, that one can impose the appropriate scaffolding on these brute facts.

Even in the case of disciplines that consists of nothing *but* principles -- purely *a priori* disciplines, such as math or philosophy – the neophyte sees those principles as brute facts. (Initially, the epsilon-delta definition of the limit is *just* a brute fact – to be memorized, not understood. Later one understands it.)

The neophyte's over-valuation of "brute facts" is reflected in the fact that, in her reasoning and discourse, she tends to rely heavily on formulas and truisms, along with the jargon in which they are embodied. There is little understanding of the concepts on which these para-truths are parasitic.

A consequence is that the neophyte's discourse tends to be very subjective. It is reflects objective logical relations holding among the data far less than it does the peculiarities of the manner in which the neophyte

assimilated that data. Instead of tracking relations of causality and entailment, the neophyte's discussion tracks facts that have little objective significance, e.g. the order in which she acquired this information and the relative degrees of emotional salience had by the data.

One characteristic of writer-based prose is that it is excessively *sequential*. It has the form: "x happened; then y happened; then z happened…" Presumably this is because facts relating to chronology can be grasped relatively easily, whereas some sophistication is needed to grasp facts relating to relations of *dependence* (causality, probabilification). A related point is that, in neophyte prose, there is a relative absence of stratification and organization. Irrelevant details are left in, and relevant meta-data is left out.

A more objective and illuminating exposition of the subject-matter at hand might have a form like:

"Here are the general facts that define the domain under investigation. At time t, x happened. Here is how x's occurrence connects with the aforementioned general facts. y happened right after x. It is still an open question whether x *caused* y to happen. Here are the considerations on both sides. Soon thereafter, z happened. It might *look* as though z happened as a result of y. But actually z came about, not because of y, but because of w. w's relation to the those background facts is thus and such…."

Instead of putting forth a heap of data that is organized, if at all, along purely chronological or psychological lines, illuminating prose derives specific data from larger principles and, in connection with this, identifies the causal and logical links underlying the raw data.

The essence of *writer*-based prose is that the organizing principles behind

it are psychological, not logical. The essence of *reader*-based prose is that those principles are logical, not psychological. The reader-based author is able to step outside herself. She is able to distinguish subjective facts relating to the manner in which she *learned* the data in question from objective facts concerning the data itself. Consequently, reader-based prose tends to be more intelligible than writer-based prose. In the latter, the data is embedded in irrelevant psychological accretions that obscure the contours of that data. In the former, those contours are in sharp relief.

But our attitude towards writer-based prose should not be punitive. In fact, it is probably necessary for first (even second and third…) drafts to be writer-based. Before the data can be organized, it must be put forth.

There are two aspects to analytic writing: first the data is identified; then the data is organized, analyzed, and processed. It is not always possible – and it is often not even advisable – for the writer to collapse both of these phases into one. If the author allows herself to state the data only on the condition that she has integrated it into the appropriate theoretical framework, then the flow of data will be stanched off and the relevant data may not make onto the page. Writer-based prose is an essential *first step* in the composition. The problem is not that writer-based writing occurs. It is that it is not always followed by the appropriate reader-based working-over of the crude data deposited on the page during the writer-based phase.

In conclusion, as writing-instructors, we must encourage writer-based prose. But we must also encourage reader-based revisions of those writer-based efforts.

Analytic Summary of Joseph Williams' "The University of Chicago."

When someone is trying to master a new subject matter, her prose will

inevitably involve errors. In fact, in its early phases, the acquisition of new information may lead to certain forms of *regression.* This is because what the author already knew has to be reconstituted within the framework defined by the knowledge now being acquired. Here is an example. Legal prose uses a number of terms not otherwise used, and it puts commonly used expressions to a new, specialized use. When a neophyte legal writer is trying to use some quasi-technical term (e.g. "whereas" or "heretofore") in its distinctively legal sense, she may momentarily forget the rules that govern its ordinary use: rules that are *still* applicable in legal writing, even though new rules have been superimposed on them.

In many cases, errors are not *just* errors, but rather reflect the fact that the author is attempting to master a new sphere of knowledge and its associated terminology. This kind of regression applies in every case where a new skill is being learned – where old methods that have worked for us so far are being replaced with superior, but initially unwieldy, techniques. So when we encounter an "error" in student writing, we have to ask ourselves whether it is *just* a slip, on the one hand, or a sign of growth, on the other. If we insist on "perfection" – freedom for error – we are likely to encourage our students to produce flawless trivialities.

Analytic Summary of Richard Haswell's "Solecism: Mistakes and errors."

The basic theme of this article is similar to that of the previous one. We are inclined to see mistakes as simple failures. But in some cases mistakes correspond to a *growth* in knowledge and competence. As you develop the ability to construct increasingly complex and information-rich sentences, you also expose yourself to greater risk of error. Suppose that you only knew a

handful of words and syntactical constructions. So you could say “the dog is hungry”, “the cat is sad”, and a few other things of this sort. Because your discourse is limited to such fatuities, it will probably never involve any syntactical errors. But the moment you enrich your prose with more complex constructions, you make error an inevitability. A Shakespeare is more likely to make errors than somebody who is telling you how to screw in a light-bulb.

When people learn new syntactic rules – e.g. those governing accusatives like “whom” and the placements of prepositions – they are likely, at least initially, to make mistakes that they didn’t previously make. This is because, in the process of mastering these new constructions, they sometimes mis-apply them or use them where they are not warranted: it takes a while to internalize this new piece of information, and errors are a part of that process. For example, somebody who learns that “whom” is to be used in cases where the antecedent pronoun is the direct or indirect object is likely to mis-apply this rather complicated rule, and thus to mis-use that word. But she is not to be scolded for this. That would only instill a fear of error that would make it impossible to ever master the new rule. So “solecisms” (grammatical or syntactical errors) aren’t always *mere* slips. Sometimes they are only *relative* slips – they occur in the context of an over-all gain in knowledge (just as a mountain-climber may falter in the context of a 20,000 foot ascent). If writing instructors make the absence of error be the main objective, then students will not do the kind of writing that benefits them – the kind where they try to move beyond what they already know – and the result will be formally correct empty shells.

The Computational Theory of Mind: Five reasons why it fails

Abstract: This paper will put forth five reasons why the Computational Theory of Mind fails. These reasons are either under-emphasized in the literature or are altogether absent from it.

0.0 Introduction According to the Computational Theory of Mind (CTM), to think is to compute. Thoughts are computations. A computation is a formal inference. So CTM is the view that any instance of thought is identical with the making of a formal inference.

But what is a "formal" inference? To qualify as "formal", a case of reasoning must be identical with a reaction, on the part of some creature or other, to symbolic-inputs. But not all such reactions qualify as cases of formal reasoning: only those reactions so qualify that aren't to any degree be driven by a knowledge of the *meanings* of the symbol-tokens in question and are driven entirely by a sensitivity to their *forms.* So CTM is the contention that thought consists of responses to symbol-tokens that are driven entirely by the formal properties of those tokens and therefore embody no knowledge of what those symbol-tokens mean. In Section 1.2, we'll give some examples of such responses, thereby eliminating any obscurity from the statement just made.

Why advocate CTM? Two reasons. First, CTM, if correct, solves the mind-body problem. When *you* add two numbers, you are *ipso facto* thinking. Computers presumably add numbers. (Computers presumably compute.) Therefore, computers, when adding numbers, are *ipso facto* thinking. We know how a computer's software is related to its hardware. So CTM, if correct, solves the mind-body problem. (In other words, CTM, if correct,

makes it clear how brains are responsible for the mental activity that they mediate.)

Second, CTM, if correct, solves the "tracking problem" (or so CTM-advocates say and so we'll momentarily assume). Brain-activity, though clearly 100% physics-driven, is also logic-driven (or, in any case, logic-*consistent*), at least to the extent that it mediates rational thought. So, to the extent that it's constitutive of rational thought, brain-activity, though governed by *de*scriptive laws, embodies a sensitivity to *pre*scriptive laws: where such activity is concerned, the descriptive "tracks" the prescriptive. Nothing that is obeying the laws of physics is, in virtue of that fact, doing anything that can be *evaluated*--- that can be described as "correct" or "valid" or "false." (In moving hither as opposed to thither, the billiard ball isn't doing anything right or wrong, valid or invalid.) And, it seems to follow, nothing that can be so described could be constituted by brain-activity. But brain-activity *does* constitute activity that can be so described.

CTM (supposedly) solves this problem. It's possible to create an environment in which there occur symbol-tokens whose logical properties are hewed to their physical properties in such a way that, within that environment, symbol-tokens occur in sequences that are consistent with the requirements of logic. In fact, not only is it possible to create such environments: we routinely *do* create them—and employ them. We call them "computers." So it would seem that CTM, if correct, explains how activity that falls within the scope of physics can, at the same time, embody a sensitivity to the demands of logic and, more generally, of normativity.

0.1 Mentalese: the "language of thought" According to CTM-advocates, those symbol-tokens mediating our thoughts are held to belong to a hypothetical language. This hypothetical language is generally referred to as

“Mentalese.”

According to CTM, one doesn’t have to *learn* Mentalese. Brains are computers. Computers don’t function until they’ve been programmed. Therefore, our brains don’t function until they’ve been programmed. Mentalese is that program. Thus, Mentalese must be programmed into a brain *before* that brain can mediate cognitive activity and, therefore, *before* that brain’s owner can learn a language or, indeed, do anything that involves having any awareness of anything. Mentalese is to a functioning brain what a computer’s operating system is to that computer. Thus, just your IBM came “bundled” with DOS, so your brain came “bundled” with Mentalese.

In light of what we’ve said, it might seem reasonable to say that, according to CTM, one innately *knows* Mentalese. But if CTM is right, one doesn’t know Mentalese at all, let alone innately. Knowing a language involves being able to correlate linguistic with non-linguistic entities. If I couldn’t correlate snow (the powdery substance) with “snow” (the word), I couldn’t possibly know what “snow” meant. In order to correlate “snow” with snow, I must have some way of being aware of snow that doesn’t involve knowing what “snow” means. Knowing Mentalese would involve being able to correlate the expressions belonging to it with non-linguistic entities. According to CTM, thoughts are identical with Mentalese expression-tokens, and such expressions must therefore be operational if one is to make correlations or, therefore, to know a language.

So if CTM is right, a *pre*condition for one's being aware of anything (or, therefore, for one's making correlations of any kind or, therefore, for one's knowing a language) is that one's brain be stocked with already operational Mentalese symbol-tokens. So it isn’t by virtue of one’s *knowing* Mentalese that expression-tokens belonging to it mediate one’s thoughts or otherwise do what CTM-advocates allege them to do.

To sum up: Nobody knows Mentalese; nobody understands the expressions belonging to it; and it isn't to any degree by being understood that such expression-tokens facilitate thought.

In this respect, Mentalese is radically different from English and every other non-hypothetical language. The expression-tokens belonging to English, Urdu, etc. belonging to such languages do not, except in a purely accidental fashion, facilitate thought *except* by being understood. By contrast, the symbol-tokens of Mentalese, supposing such tokens to exist, need not be understood to facilitate thought. Indeed, it is inherent what Mentalese expression-tokens are that they be able to discharge *all* of their duties *without* being understood to any degree by anyone or anything.

Because Mentalese, were it to exist, would differ in this respect from English, Spanish, and other *bona fide* languages, it follows (for reasons to be discussed in Section 4.0) that the very concept of a "language of thought" is not a coherent one. Given that CTM assumes there to be a language of thought, and given that (for reasons to be disclosed in Section 4.0) no doctrine that doesn't assume this bears any resemblance to CTM, it follows that CTM is incoherent

0.2 The concept of a formal inference: some illustrations We've seen, *if* correct, CTM explains much. But *is* it correct? We can't answer that unless we make it clear what, in this context, the word "compute" means. And we can't make *that* clear unless we make it clear what, in this context, it means to say of symbolic activity that it is "form-driven."

The following three stories will help answer these questions.

Story #1 (*not* an example of computation) Smith, who speaks English, is told: "Jones has two yachts." On the basis of this information, along with his knowledge of the English language, Smith concludes that the following

sentence is correct: “the number n such that Jones has n yachts is such that n has no rational square root.”

Commentary on Story #1: Smith’s reasoning is cogent. But it is not in any sense “formal.” Smith’s thought process was a driven by an understanding of *inter alia* the *meanings* of the words he was hearing.

Story #2 (an example of *a* kind of computation) W is a wood-chipper. But W is no ordinary wood-chipper. Suppose that W is fed, in the following order, a wooden “2”, a wooden “+”,a wooden “9”, and, finally, a wooden “=.” Suppose, in other words, that W is fed an instance of the sentence=fragment “2+9=.” In that case, driven *solely* by the morphological properties of the symbols that it was fed, W cranks out a wooden “11.” And, W does the same thing *mutatis mutandis* for any other two numbers. So given any whole numbers M and N, if W is fed the sentence-fragment “M+N=”, then, driven solely the morphologies of that fragment, W produces a wooden sign denoting the sum of those two numbers.

Commentary: There would seem to be *a* sense in which W can compute sums. It is typically *this* sense of the word “compute” that advocates of CTM have in mind when they say that to think is to compute. Why only “typically”? Why not “always”? Because sometimes what they have in mind is the *prima facie* quite different sense of the word “compute” that is illustrated by:

Story #3 (an example of a different kind of computation) Smith has no idea what “1,245”, “241”, “1,486”, etc. mean. But Smith has been told by a source whose testimony he rightly trusts that, when given ink-marks having

such and such shapes, a correct sentence results when another ink-mark, having thus and shape, is placed in a certain position relative to the original ink-marks. In particular, Smith knows that, given their respective shapes, the ink-marks “1,486” and “1,245+241=” form a correct sentence when (some isomorph of) the former is placed to the right of (some isomorph of) the latter. It must be emphasized that Smith has no idea what any of these ink-marks means. What Smith *does* know—and this is *all* he knows (that is of relevance of the present context)---are the rules of adding multi-digit numerals *so far as those rules can be identified with the rules concerning the manipulation of shapes*.

Commentary: What Smith is doing seems to be different from what W is doing. Smith’s applying the just-mentioned rule involves his having various sense-perceptions (e.g. he sees various ink-deposits) and various beliefs (e.g. he believes that those ink-marks are symbols of some kind or other, and he believes that a correct statement---albeit one of whose meaning he is ignorant—will result if, underneath them, he produces an ink-mark having a certain shape); and his having such beliefs involves his making quite delicate judgments (e.g. he must judge that the ink-deposits in question were produced by sentient beings with certain intentions---he would *not* see those ink-deposits as tokening the sentence-fragment “1,245+241=” were he to learn that their existence was the unintended result of some explosion involving an ink-filled urn).

Nothing comparable holds of W. It isn’t because of any sense-perceptions or beliefs or intentions that it has that, having been fed a wooden token of the sentence-fragment “2+9=”, W spits out a wooden “11.” So what Smith is doing in writing a “1,486” seems *not* to be comparable to what W is doing in spitting out a wooden “11.”

And, while there is certainly *a* sense of the word "compute" in which Smith is computing a sum, this cannot possibly be the sense in which, according to CTM, to think is to compute. A *pre*condition for Smith's performing the just-described computation is his *already* being able to upload and process information. So CTM would be guilty of vicious circularity were it given by the thesis that, in *this* sense of the word "compute", to think is to compute. So, supposing CTM coherent, the sense in which to think is to compute is the sense in which W computes, and the sense in which *Smith* computes is irrelevant.

We will now see that CTM is indefensible and that no doctrine at all similar to it can possibly be correct. There are five reasons for this. These reasons will given insections 1.0-5.0.

1.0 CTM's failure to distinguish between shape and syntactic form A story will help expose one reason why CTM is false. You shout the following at me: "if your nose starts to bleed in the next ten seconds, I'll give you a million bucks." Let T be the sentence-token in question. As a result of T's being uttered, I *do* start to bleed out of my noise, and you duly give me $1,000,000. But it wasn't because of what T meant that I had a nose-bleed; nor was it because of T's syntax. It was T's being loud that caused my nose-bleed.

My reaction here wasn't *symbolic.* Even though T is a symbol (or, to be precise, a symbol-token) T's being a symbol did nothing. For exactly similar reasons, W's spitting out an "11" isn't symbolic---even though its doing so was precipitated by events involving symbol-tokens.

1.1 This line of reasoning generalized We've just seen that CTM is non-starter if a symbol-token's "form" is identified with its shape. But sometimes,

when referring to a symbol-token's "form", advocates of CTM are (so they say) referring to its *syntactic* form. Given this use of the word "form", CTM becomes the contention that thoughts are syntax-driven operations on symbols.

This contention is easily refuted. Syntax is semantic structure. "Smith loves Mary" and "Sally hates Greg" have the same syntax because the semantics of the one has the same *structure* as the semantics of the other.

To be sure, those sentences have very different meanings. But they're semantically isomorphic. Each constituent of the one sentence corresponds to exactly one constituent of the other ("Smith", "loves", and Mary" correspond to "Sally", "hates", and "Greg", respectively); and the relationship borne by the semantics of any given constituent of the one sentence to the semantics of that sentence as a whole coincides with the relationship borne by the semantics of that constituent's counterpart in the other sentence to the semantics of *that* sentence as a whole. So the *relation* borne by the semantics of "Smith" with respect to the proposition that Smith loves Mary coincides with the relation borne by "Sally" to the proposition that Sally hates Greg.

Given that syntax is semantic structure, syntax-driven operations are *ipso facto* meaning-driven. Being the contention that "formal" operations are *not* meaning-driven, CTM is therefore completely incoherent if taken to be the contention that "formal" operations are syntax-driven operations.

At the same time, CTM is incoherent, as we saw, if "formal" operations are taken to be shape-driven operations. There is no third way to interpret the word "formal." So CTM is incoherent.

1.2 Shape vs. syntax (revisited): the equivocation that lies at the heart of CTM

Here is the thinking embodied in CTM:

"Form-driven responses to symbolic inputs are *ipso facto* cases of thought. W has form-driven responses to symbolic inputs. Therefore, W is thinking (it is adding numbers."

But when this argument is stated in a clear and non-elliptical manner, it becomes clear that it involves an equivocation and is invalid for that reason:

"Form-driven responses to symbolic inputs are *ipso facto* cases of thought. (In other words, *syntax-driven* responses to symbol-tokens are *ipso facto* cases of thought.) W has form-driven responses to symbolic inputs. (In other words, W's responses to its inputs are *morphology*-driven.) Therefore W is thinking."

The second argument is clearly spurious. Presumably because they are aware of this, advocates of CTM typically hold that, even though a symbol-token's shape isn't *identical* with its morphology, syntax can be "coded" into morphology---that syntax can be "morphologized" in much the way that geometry can be arithmetized. This is both false and irrelevant.

It's false because anything, no matter what it's shape, can have any syntactic structure, and anything, no matter what it's syntactic structure, can be tokened by anything, no matter what its shape. (In other words, given an arbitrary shape, and given an arbitrary syntactic structure, there is a possible language to which there belong expressions having that syntactic structure that are tokened by objects having that shape.) It's irrelevant because, even if syntax can be morphologized, it isn't, as we've already seen, *in virtue of* its having this as opposed to that shape that a symbol-token has this as opposed

to that syntax. So *in so far as* one is responding to an object's shape, one's behavior is, to that extent, *not* the effect of, or otherwise to be understood in terms of, that object's shape.

To be sure, one knows what syntax an object has *on the basis of* its shape. A large part of the reason that I know the syntax of the following sentence-token---"John is tall and grass is green"---is that I know what shape that ink-deposit has. But sentence-tokens with precisely the same syntax can be tokened by hand-movements *or* by noises; and hand-movements and noises have no strictly morphological resemblance at all to any ink-deposit. (No ink-deposit that tokens "John is tall and grass is green" bears much resemblance to a stream of noises, or a series of hand-movements, that also tokened that same sentence, except relative to rules of projection that were contrived *ex post facto* to establish such a resemblance.)

Also, the relationship that holds between (a) my knowledge of that ink-deposit's shape and (b) my knowledge of its syntax is of a *causal*, and not of a *constitutive*, nature. I see that it has that shape. Given that knowledge, along with my background linguistic knowledge, I judge that it has a certain syntax.

To be sure, from an experiential (phenomenological) perspective, it may seem as though *in* knowing that it has that shape I *ipso facto* know that it has that syntax. But that's not the case. What we have is one bit of knowledge (knowledge of shape) *leading*, by way of (rapid) judgments, to some other bit of knowledge (knowledge of syntax).

Echoing what we just said, these two bits of knowledge may seem, at the level of consciousness, to come into existence simultaneously. But that appearance is misleading and, even if not misleading, irrelevant. It's misleading because any knowledge of an object's syntax *must* follow knowledge of its physical properties. To clarify these last two statements: If I judge that you're angry, it's on the basis of your facial expression, or some

such. My judgment as to your emotional state may immediately follow my awareness of your expression. But given that the former is *based* on the latter and is also *caused* by it, the one awareness follows the other, however quickly. Exactly similar points hold of the relationship between my knowledge of a given object's physical and my subsequent knowledge of its syntax: it isn't possible for me to have the latter knowledge without having the former; it *is* possible for me to have the former without having the latter; and my knowledge of the former is the sole foundation for, and is also what leads to, my knowledge of the latter.

And it's irrelevant that, speaking from a strictly experiential (phenomenological) perspective, the occurrence of the one bit of knowledge *seems* to coincide with the occurrence of the other. It is the purpose, or at the very least the effect, of consciousness to condense the intricacies of the cognitions underlying it into a single, seemingly homogeneous "experience." If every articulation of pre-conscious cognition were faithfully represented in consciousness, if phenomenology faithfully duplicated every joint in the intricacies of the cognitions of which our conscious judgments are the outcomes, then consciousness would be so overburdened with information that it would cease to be able to perform the executive functions with which it is tasked. When seeing a rabid dog run towards me, my consciousness would present to me a richly articulated mosaic of judgments, whose consequences for my well-being would take much more time to figure out than was available to me. Fortunately, however, my consciousness presents me with a homogenous feeling that causes me,not a moment too soon, to bolt into action. Instead of giving me the psychological equivalent of a two-hundred page state-department report, consciousness gives me the equivalent of an electric cattle-prod in my back. And because the rich cognitive back-story is telescoped into that bolt of unpleasant and, aside from its distinctive affective

nuances, homogeneous bolt of unpleasant phenomenology, I do the right thing (i.e. I run to the mini-van, lock the doors, roll up the windows, etc.) and, what is more, I do it in real time.

In general, it is precisely *because* consciousness represents as homogenous "feelings" what are in fact intricate networks of judgments that the contents of consciousness, unlike their cognitive underpinnings, are *actionable.* It is, I would conjecture, the very purpose, or at least one of the purposes, of consciousness to condense cognitive intricacies into affective simplicities. Therefore, it is *inter alia* its purpose to condense distinct, albeit adjacent, stages of intricate thought processes into a single, action-causing, survival-ensuring *feeling*.

In any case, if this is at all on the mark, then it's of no relevance that my recognition of the ink-deposit's shape *seems* to be temporally coincident with my recognition of its syntax. –It is to be pointed out that, if this *isn't* on the mark, then much work in psychology crumbles.

1.3 A related point: CTM incompatible with the fact that distinct thoughts can be part of some one thought-process If CTM is right, my thinking *2 is a prime number and 2 is less than three* cannot be responsible for my later thinking *some prime number is less than three.* According to CTM, if B1 is the totality of brain-events constitutive of my thinking *2 is a prime number and 2 is less than three*, and B2 is the totality of brain-events constitutive of my thinking *some prime number is less than three*, it cannot be B1's *being* such a belief that brings about B2. So it cannot be B1's being such a belief that leads me to believe that some prime number is less than three.

So, if CTM is right, B1's being such a belief has nothing to do with B2's existence or *a fortiori* with B2's being constitutive of a thought to the effect

that some prime number is less than three. Thus, my believing that two is even and is less than three cannot, if CTM is right, have anything to do with my later believing that some even prime is less than three. So, if CTM is right, T1 has nothing to do with T2, notwithstanding that the brain-events constitutive of T1 brought about those constitutive of T2.

So even though B1 occurs right before B2 and is causally responsible for it, and even though B1 and B2 are constitutive of, respectively, T1 and T2, nonetheless T1 and T2 cannot, if CTM is right, be installments in some one thought-*process.*

In general, CTM is incompatible with fact that distinct thoughts can be part of some one thought-process. CTM is thus incompatible with the fact that thought-processes occur---with the fact that there are *sequences* of thoughts (setting aside such thought-sequences whose constituents have no thematic coherence, having at best the content-free coherence had any, e.g. a series of domino-fallings).

2.0 CTM's dependence on the false assumption that syntax can mirror logical form CTM assumes that, under some conditions, morphology-driven activity is *ipso facto* syntax-driven. We've found this assumption to be false. But in this section we will grant it for the sake of argument.

Like all other natural languages, the English language is replete with context-sensitive expressions. And were it not for that fact, its expressive capabilities would be pitifully limited. There's a big difference, semantically and prudentially, between saying (truthfully, we'll suppose)

(i) "that obviously rabid animal will soon start heading in our direction, and we'll be eaten alive if we don't immediately get in that guy's mini-van"

and saying (same qualification)

(ii) “at time t, a certain rabid animal r is in place p and at time t*, which is soon after t, x will start rolling towards place p*; and unless, during the interval between t* and when said animal arrives at p*, individuals x and y get into person z’s mini-van, x and y will be eaten alive.”

even if the values t, p, and p* correspond to times and places indicated by somebody who uttered (i). Let’s assume that, in (i), the referent of “that obviously rabid animal” *does* in fact pick out a rabid and (for that reason, among others, dangerous) animal.

Let T be a token of (i), and let T* be a token of (ii). It’s clear that, pretty much no matter what the circumstances are, the information borne by T is different—semantically and prudentially—from the information borne by T*.

For this reason, and countless others of its ilk, a language that didn’t contain expressions like “that”, “here”, “I’, “now”---context-sensitive (indexical) expressions—would be quite useless (for everything other than, suggestively, making time- and place-insensitive statements about numbers and other denizens of Plato’s heaven).

Bearing this in mind, let T# be a token of the same sentence-*type* as (i). But let’s assume that the occurrence, in T#, of “that obviously rabid animal” refers to an animal that, though rabid, is one tenth the size of a shrew, and also has no teeth and thus no way of biting (or otherwise transmitting rabies to) anyone, and is therefore no threat at all. In that case, clearly, T# is false, whereas T (so we may suppose) is correct.

T and T* have the same syntax—on any delineation of that term. (Same grammar; same depth-grammar; same logical form.) But the proposition meant by T has very different entailment-relations from the proposition

meant by T*. And, what is more on point, a creature that responded only to the *forms* of *sentences* would (other things being equal) respond to T in the same way that it would respond to T*. But (other things being equal) no person would respond to T in the same that he'd respond to T*. In general, beings that respond to sentence-token-form, and not (except in so far as sentence-form *is* (low-grade) sentence-semantics) to sentence-token-semantics, are (i) not rational creatures and (ii) aren't us (since, irrational though we often are, we *tend* to be irrational---we flee from tigers, but not from pebbles).

Replace T and T* with their Mentalese counterparts. In other words, take them to be tokens, not of English, but of Mentalese sentences. It is evident that the resulting argument goes through, showing that we don't think in Mentalese.

2.1 An objection to an objection to this line of thought "There's a problem with your argument", it might be protested. "What if Mentalese is such that the context-sensitivity that, in English (and Spanish etc.) is accomplished through context-sensitive expressions, is accomplished by strictly syntactico-logico-formal means? In that case, the argument you just gave would lose whatever anti-CTM value it has."

This suggestion is a non-starter. Where there is context–sensitivity, there is *ipso facto* a cleavage between sentential (grammatical) and logical (propositional) form. Being inherent in the nature of context-sensetivity, this fact nothing to do with the peculiarities of English or any other language. A context-sensitive expression is one whose *referent* varies from context to context. Reference is a quintessentially semantic function. So there is no way to de-semanticize context-sensitivity.

So if we think in Mentalese and, and if, moreover, Mentalese, like English,

contains context-sensitive expressions, then our mental activity is semantics-driven and consequently isn't computational, an immediate consequence being that CTM is false. Human thought is context-sensitive. It follows that, if we think in a language, that language *must* be context-sensitive, and therefore semantics-driven. So if we think in a language, CTM is wrong. And CTM is wrong if we *don't* think in a language, given that, according to CTM, thought consists entirely of formal operations on linguistic symbols.

3.0 CTM's reliance on the erroneous assumption that there are minimal units of thought First a definition: a "morpheme" is an expression that cannot be decomposed into others. Bearing this in mine, let's suppose that, just as CTM says, we think in a language. In that case, our concepts are identical with expressions. Linguistic expressions consist of morphemes. (Any given complex expression (e.g. "the cat is on the mat") decomposes into expressions that cannot, in their turn, be further decomposed (e.g. "cat", "hat"). So in contending that our thoughts are identical with (tokens of) Mentalese symbols, CTM is, in effect, assuming that we have simple concepts: concepts that cannot be decomposed in others. So CTM, if correct, entails that any given thought that one has can be decomposed into "simple" concepts. CTM thus has the radically implausible consequence that we have "simple" concepts. In other words, CTM has the consequence that some of our are no more capable of being decomposed into other concepts than a morpheme (e.g. the word "cat") can be decomposed into other expressions. If this consequence of CTM is correct, it follows, as we'll now see, that some of our concepts would be such that one could have them without also having other concepts.

Let T be the token of the Mentalese word for "snow" that went off when, just a moment ago, I thought to myself that snow is white. We may assume

that T is a morpheme. (If, for some reason, you reject that assumption, just map what we say about T onto some Mentalese expression-token that, in your judgment, is a morpheme—and you must, if you grant that we think in a language, grant that there is some such token, since morpheme-less language is a surd---and everything that we say about T will hold *mutatis mutandis* of this other expression-token.) Not being decomposable into other expression-token, T presumably exists independently of such tokens. Given any *simple* expression—given any morpheme—belonging to a language L, you can remove that morpheme without removing other morphemes. Of course, in removing that morpheme, you'll remove all of the expressions of which it is a constituent. (And, if you choose a non-lexical item, e.g. "and" or "of", you may thus do a lot of damage to L.) But you won't, in removing that morpheme, be removing *other* morphemes.

Therefore, if, as CTM alleges, we think in a language ("Mentalese") , some of our concepts coincide with morphemes of a language, if follows (i) that if you strip Mentalese of the word-type of which T is a token, the *only* other expressions which, in so doing, you'll be removing from Mentalese will be other morphemes, and it also follows (ii) that some of our concepts are such that one can have them without having others. This means that, if CTM is right, your concept of snow is *primitive*, the same being true of your concept of Smith, Jones, and Brown (assuming, for argument's sake, that the Mentalese expressions for these individuals are morphemic).

3.1 CTM's reliance on the erroneous assumption that there are minimal units of thought (continued) But there are no primitive concepts. I couldn't possibly have a concept of snow, or of Barack Obama, or (to take a more tendentious example) of the number three without having a network of other concepts---extension, spatiotemporal location, surface-area, sentience, etc.

Concepts are necessarily contrastive; in order to have a concept of a thing, one has to be able to distinguish that thing from other things. A pre-requisite for distinguishing, say, that rock from, say, that tree is my being aware of some property had by the rock that isn't had by the tree. And a prerequisite for my having such knowledge is my *already* having concepts. (Many contemporary theorists---some of them CTM-advocates, some of them not—reject this, saying that conception is a purely causal notion. Their view is that concepts are mental photographs. The relationship borne by my concept of Socrates to the Socrates is identical with, or to be understood in terms of, in terms of, the relationship borne by a photograph of Barack Obama to the events (the impact of certain light-rays with Barack Obama's body) that, in due course, led to that photograph.

This view, as we've seen, is false; and so are the modifications of put forth by CTM-advocates in their attempts to sidestep the glaring problems with it.

What right have I to make such an encompassing claim? Every right. Here's the idea that I am now claiming to be false and all of whose derivatives I am now claiming to be false:

(PM) My concept of Socrates is to Socrates what a photograph of Socrates is to Socrates.

That idea involves a failure to distinguish between (i) awareness of some entity and (ii) information that, *if known*, would *lead* to such awareness. I see discolorations of a certain kind on a photographic plate. Given that information, along with other pieces of background information that I have, I can make various inferences. I can infer that, at some point, light bounced off of a person who was bald and was wearing a toga (and so on). So the photographic plate is information in the sense that, to a creature with the

requisite background and the requisite intelligence, its existence warrants the making of an inference that will eventuate in knowledge of some entity (e.g. a person) or state of affairs (e.g. one consisting of a person wearing a toga) that might otherwise have remained unknown. (Why "*might* otherwise have been unknown", instead of "*would* otherwise have been unknown"? Because I might have been the one who took the picture and might therefore have already have known of the person and the corresponding state of affairs.) The discolorations on the photographic plate are "information" in the sense that, to a suitably endowed creature, they *warrant* the making of inferences that, in due course, lead to concepts.

But the photograph *itself* obviously isn't a concept of Socrates---or of anything else. And the "photograph-model" of conception, to use Gareth Evans' term, involves the grotesquely mistaken idea that concepts *per se* are identical with the evidences on the basis of which concepts are arrived at. There have been many attempts to refine the photograph model. But, in addition to being both *ad hoc* and devoid of intuitive force, they all involve this exact blunder; and, suggestively, the expositors of these views (e.g. Fodor (1990, 1998, 2006), Dretske (1982)) never so much as mention the distinction between (i) the concept and (ii) the information on the basis of which one generates the concept.

In any case, it is massively to CTM's discredit that it presupposes that there are representational atoms.

3.2 A similar argument Sentences are digital structures. Setting aside cases of ambiguity, any given sentence—any given complex expression, in fact---has a unique decomposition into minimal units of significance. This means, as previously pointed out, that our thoughts decompose into minimal representational units. Which is radically implausible.

In any case, sense-perceptions don't have a digital structure, and neither, it would seem to follow, does the information that sense-perceptions encode. Given which, it would be very artificial to maintain that all post-perceptual cognition was digital or, therefore, sentential in nature.

But, for argument's sake, let's suppose that the contents of our sense-perceptions can, as come advocates of CTM maintain, be "digitized." Supposing this hypothesis correct, all the problems previously posed for CTM by perceptual content are now posed by the post-perceptual thoughts involved in this (alleged) digitization-process. If I'm translating from Albanian to English, I must speak both Albanian and English. Moreover, supposing that I'm translating competently, I have some way of evaluating the degree to which my English outputs match my Albanian inputs. If my thinking is entirely mediated by English sentences, I won't be able to do either of these things: I won't be able to translate from Albanian into English; and, if *per impossibile* I were able to do so, I couldn't verify the accuracy of my translations. By analogous reasoning, if the aforementioned post-perceptual thoughts were themselves entirely digital, they would *ipso facto* be blind to the benchmarks relative to which their own translations were to be evaluated, supposing for argument's sake that they could produce such translations to begin with.

Of course, one could say that those processes are purely causal, i.e. that they don't embody any awareness of what, given the relevant perceptual inputs, the relevant digital outputs *ought* to be. But in that case, our post-perceptual thoughts are *ex hypothesi* merely the *results of*, and are in no way *justified by*, the contents of our perceptions. But that is surely too absurd a position to take seriously.

4.0 The regress-problem If CTM is right, then ("AS" is short for "for

anything representational symbolic”

(ARS) anything that represents anything is a symbol.

If, as CTM holds, we think in sentences—if, as CTM says, our thoughts are sentence-tokens---then surely we understand those tokens. But understanding a sentence involves being able to pair it off with its meaning. Which means that one must be able to compare that sentence with something that isn’t a sentence. Which wouldn’t be possible if all of our thoughts were sentences (or sentence-tokens).

“That doesn’t follow”, it might be said. “Suppose that you think that snow is white, and that your thinking this consists in there being an occurrence of some Mentalese sentence. Your understanding that sentenced could consist in there occurring, in your mind-brain, various sentences of some language—call it Mentalese2---such that, given those sentence-occurrences, you can duly interpret the original (Mentalese) sentence.”

Yes, but by that logic, there must be tokens of Mentalese3 sentences before one can understand those Mentalese2 sentence-occurrences; and there , there must be tokens of Mentalese4 sentences before one can understand those Mentalese3 sentence-occurrences; and so on *ad infinitum.*

So it’s viciously regressive to hold, as CTM does, that symbol-tokens mediate our thoughts. More precisely, it’s viciously regressive *given* the reasonable assumption that we *understand* those symbols.

4.1 How advocates of CTM deal with the regress-problem Aware of this line of thought, advocates of CTM say that we *don’t* understand the symbol-tokens that mediate our own thoughts. As William Lycan put it, we are simply “built to use those symbols” in the “right” ways, much as computers

are “built” to use the symbols (or, rather, the things which are tendentiously and, we have seen, wrongly *called* “symbols”) on which it operates.

But, so far as a given thing isn’t understood, it isn’t functioning *as* a symbol. A story, similar to one told earlier, will make this clear. I have a wooden sculpture—which we’ll call “WS”---on my desk that tokens the sentence “haste makes waste.” Even though WS tokens three words, all three word-tokens are carved from the same block of wood. So WS consists of one, solid object, not of a a lot of letter- or word-blocks. WS heavy; it’s made of solid oak, not pine or plywood or any other cheap sort of wood. It would thus serve as an excellent cudgel. A student of mine comes into my office. He reads it, understands it, and accordingly changes his, hitherto impetuous ways. In this context, WS is functioning as a symbol. In other words, if E1…En are the just-mentioned effects on the student, it is WS’s being a symbol (of quite a specific kind, in fact) that leads to E1…En. Later that afternoon, a colleague, whom I detest, come into my office. He doesn’t read WS. Because of his myopia, he sees a brown blur in the place that WS occupies. Gripped by a non-tenure-conducive rage, I strike my colleague with WS. If E1*…Em* are the subsequent bruises and fractures suffered by my colleague, it is WS’s having a certain mass, moving with a certain velocity, etc. that brings about E1….Em*. WS’s *being a symbol* is utterly irrelevant. Why? Because WS’s being understood is essential to its functioning as a symbol.

4.1.1 This line of reasoning summed up Languages *are* sets of semantic rules. Where there no such rules, there are no languages; and where there is no knowledge of such rules, there is no linguistic activity. The idea that there is a “language of thought” is incoherent, given that it is inconsistent with these basic truths, and so is CTM, given that it assumes that there is a language of thought.

5.0 CTM self-refuting. CTM says that we think in sentence-tokens. CTM also says that, *qua* proposition-bearing entities, those sentence-tokens *do nothing*. For reasons that we'll now see, it follows that, if CTM is correct, it cannot be *known* to be correct or even rationally accepted.

We must start by remembering that no *thing* is a cause. The rock isn't what breaks the window. What does so is the rock's striking the window with a certain amount of force. Bearing this in mind, let B be the brain-structure that encodes my belief that 1+1=2. Nobody denies that *B* has causal properties. But if CTM is right, B's being (or constituting) a thought to the effect that 1+1=2 does nothing. The same thing *mutatis mutandis* holds of any content-bearing brain-state. If CTM is right, even if a given brain-state bears content C, it's bearing that content is causally inert. This entails that, if a given person accepts CTM, it isn't because of the contents of any of the thoughts that he does so; and, consequently, it isn't because he knows of any reason to accept CTM. So, if CTM is correct, there is *ipso facto* no way that a person could be justified in accepting CTM (or, by obvious extensions of this reasoning, anything else). Which, while not showing that it's false, does show that one cannot coherently advocate it.

Berkeley and Contemporary Anti-foundationalism[257]

Abstract: Wilfred Sellars destroyed the 'myth of the given' and thus began a revolution in epistemology. Foundationalism used to be the orthodoxy. Nowadays, because of Sellars, *anti*-foundationalism is the orthodoxy. This paper shows that, two hundred years before Sellars, George Berkeley

destroyed the myth of the given by means of an argument very similar to Sellars'. Berkeley's argument is actually a kind of generalization of Sellars' argument. Berkeley explicitly shows how anti-foundationalism applies to a wide range of contexts. Sellars, by contrast, discusses only very special contexts, and leaves it entirely implicit how his discussion ought to be generalized. Also, Berkeley provides a non-metaphorical way of interpreting some of the metaphors that dominate contemporary anti-foundationalism (e.g. the idea that one's belief system is an "arc" or a "web").

Here is one conception of how we learn about external reality:

"No empirical background knowledge is needed to *obtain* knowledge of external reality. A given perception can do that on its own. All empirical knowledge rests on top of perceptions. So, in their turn, perceptions cannot, and therefore do not, rest on top of empirical knowledge. Therefore perceptions provide us with knowledge of the external world without the assistance of any *background* knowledge."[258]

Wilfred Sellars demolished this picture. In so doing, he destroyed 'the myth of the given'. It must be made clear what the word 'given' means here. Obviously Sellars isn't denying that external realities are given to us, i.e. made to known to us, through sense-perceptions. No one would seriously deny that. Sellars is attacking the idea that, when a perception makes us aware of some external state of affairs, it does so *without* the cooperation of a manifold of preexisting empirical knowledge. So he is against the idea that

we don't have to deploy existing empirical knowledge to have a perceptual awareness of external reality.

So there are *two* senses in which, for Sellars, knowledge acquired through a perception is not 'given'. It is not a given in the sense its existence typically *presupposes* a background of empirical knowledge. Background knowledge must be mobilized if a sensory experience is to give us knowledge of an external state of affairs. What is given – the sensory experience, considered apart from the background knowledge with which it interacts - is not epistemic, i.e. it does not constitute *knowledge* of anything. So information about the external world lies in what is given to us *plus* a pre-existing manifold of knowledge; such information is not located in what is given *per se*. Thus the given is not epistemic.

But Sellars is also making a stronger claim. The significance of a quale – what it truly or falsely *tells* the subject, leaving aside what the subject knows in virtue of having it – is not *given*; it is not fixed, at least not by the intrinsic properties of the quale. The quale has to be filtered through background knowledge first: *then* the quale gives the subject a determinate message about the external world. So the epistemic significance of a quale is not fixed or 'given'; it is a variable function of background empirical knowledge. Sellars is against the idea that 'raw sensations' (as he puts it) *by themselves* have a fixed or 'given' message and, thus, don't need the co-operation of ulterior knowledge to tell us anything, whether true or false, about the outside world. *A fortiori* Sellars is also against the idea that, by themselves, raw sensations constitute *knowledge* of the external world.

So there are two myths of the given. (i) What is given is epistemic, i.e. constitutes knowledge. (ii) What is given is representational. i.e. it tells us *anything*, whether true or false. Sellars made a compelling case against both myths.

It should be made clear that, according to Sellars, what is false is not the belief that there is a given.[259] Obviously the perceiver is passive in *some* respects; *some* things are given to him. What is false is the belief that the given is epistemic, i.e. constitutes knowledge, and also (more radically) that what is given is even representational.

I believe that, in *An Essay Towards a New Theory of Vision* (1709), George Berkeley made essentially the same points that Sellars would make, over two hundred years later, in *Empiricism and the Philosophy of Mind.* Further Berkeley's attack on the given is more perspicuous, more radical, and more cogent than Sellars' version of it.

Sellars argument is basically this. Bob is in a tie shop where there is unusual lighting. Bob is looking at a tie. Because the lighting conditions are unusual, he cannot tell whether he is looking at a green tie or a blue tie. So he takes the tie outside, where the lighting conditions are 'standard', and Bob sees that it is (say) blue. Bob then goes back inside and realizes what a blue tie looks like under the lighting conditions that prevail there. He repeats this procedure with ties of many different colors. So, in due course, Bob doesn't *have* to leave the shop to tell what a color a tie is; he can just *see* that it is blue or green – even though a blue tie in the tie shop gives off the same appearance as a green tie under more standard lighting conditions.

Let Post-Bob be our word for Bob *after* he has 'acclimatized' to the unusual lighting conditions in the tie-shop. And let Pre-Bob be our term for Bob before that acclimatization. Confronted with a tie in the tie-shop, Post-Bob can *see* that it is blue or green – even though it doesn't *look* the way that blue or green ties look ordinarily. But obviously when Post-Bob *sees* what color a tie is, a great deal of back-ground empirical knowledge is at work. He knows a great deal about the conditions of perception that obtain in the tie-shop, and how they affect the appearance of the ties. Of course, if Post-Bob is

to ascertain what color a given tie is, he will simply look at it. But from this perception, Post-Bob is able to upload the color of the tie only because, operating in his cognitive background, there is a fair amount of *existing* empirical knowledge.

There are two ways that this lesson can be generalized. (Sellars leaves implicit these much needed generalizations.) It can be shown to apply to sense-perception under *standard* conditions (i.e. under conditions where the lighting, and so on, is relatively 'normal'). It can also be shown to apply to sensory apprehensions of properties *other* than colors.

Consider the highly indirect way that Post-Bob ascertains the color of the tie (in the tie-shop). It would be possible to construct scenarios in which someone ascertains knowledge of *other* properties of physical objects in a similarly indirect way – in a way that involves the co-operation of background knowledge (concerning the conditions of perception). We could imagine Bob being in a place where – because of, say, the distorting effects of funhouse-style mirrors – Bob could not (at least not initially) *see* what shapes objects were. He'd have to look at a given object under 'standard' conditions to see what shape it was, and then look at it again under the unusual conditions just described: he would *then* know what (say) a square or triangle shaped object looked like under *those* circumstances. Of course, eventually, Bob would just be able to *look* at an object (in the place where the 'bent' conditions obtain) and, without further ado, know what shape it had: in effect, he would be able to *see* that it was a square. But whenever he did this, it would be, in part, because of the co-operation of background empirical knowledge.

Given some ingenuity, it would be possible to construct scenarios that do for mass, temperature, speed, distance, or size what the scenarios we've just considered do for color or shape.

What Sellars says about Post-Bob's way of ascertaining the colors of ties applies to the way that *we* obtain knowledge of objects' (perceptible) properties under 'standard' conditions. In other words, when *we* are just *looking* at an object and seeing that it is red, the perception therein involved is informed by background empirical knowledge in just the way that Post-Bob's perception of the color of a tie is thus informed.

First of all, the conditions of perception under which we in fact sense-perceive objects are in a constant state of flux. In some cases those changes are quite dramatic. But unless the changes are *highly* unusual, we adjust automatically and unthinkingly. You simply *see* that (say) some tie is red in circumstances where the lighting conditions are quite non-standard (unless, of course, the conditions are *exceptionally* non-standard). Unless you are an artist, you are probably unaware how massive these differences are. For example, my apartment has excellent natural lighting. But my artificial lighting – the lighting I use after night-fall -- is extremely poor. Thus, the lighting conditions in my apartment at 3:30 p.m. are very different from the lighting conditions in my apartment at 10:00 p.m. (with the lights turned on). The way that a red book appears is *in fact* very different depending on the time. This became glaringly obvious when I took up painting again. The mix of paint that I must use to paint the book as it appears at 10:00 p.m. is very different from the mix I must use to paint the same book as it appears at 3:00 p.m. or even 8:30 p.m.

But if I want to find out what color a book is, I don't have to go through the rigmarole that Pre-Bob went through: I just look at the book. But does this mean that background knowledge is not *in fact* a factor in my ascertaining the color of the book? It seems much more natural to suppose that such knowledge *is* operative, but that I have so thoroughly internalized the correspondences in question – the correspondences that Post-Bob had to

work *so hard* to learn, given the extreme unusualness of the lighting conditions in the tie-shop – that I am no longer aware of the role that this ulterior empirical knowledge plays.

Sellars has shown that in non-standard conditions, one must rely (however unconsciously) on background knowledge if one is to learn, through one's senses, how things are in the external world. But standard conditions are an idealization: *no* conditions are really standard. If there were perfectly 'standard' conditions, they would not be standard. If we could set up a situation that was distortion free, that situation would *not* be a typical one. So what Sellars says about Bob's intake of color-facts in the tie-shop is true of our intake of color-facts *in general.*

Further, it is true of our intake of shape-facts, size-facts, distance-facts, and so on. We can tell at a glance what shape an object has. But the appearance that a diamond-shaped object gives off when looked at from above is the same as the appearance that a square shaped object gives off when looked at from an oblique angle. But we *instantly* know, except in contrived circumstances, whether an object is diamond-shaped or square shaped. We *see* that the object is (say) square. How can we *see* this, given that its *appearance,* considered by itself, could just as well be that of a diamond-shaped object? Because we take into account a great deal of contextual information; in particular, we take into account the fact that we are not hovering above the object.

A ten foot stop-sign, looked at from a hundred feet away, gives off the same appearance as 1,000 foot stop sign looked at from 10,000 feet away. But when one looks at a stop sign from a hundred feet away, one isn't fooled for a second; one *sees* that it is a stop sign (of a certain approximate height); one isn't fooled for a moment into thinking that it is 1000 ft stop-sign. (Of course, one cannot tell, just by looking at a stop-sign, *exactly* what height it

is. But one is seldom off by orders of magnitude.) One can just *see* the approximate height of the stop sign. But how is this possible, given that a ten foot stop sign looked at from a hundred feet away gives off the same appearance as 1,000 ft stop-sign looked at from 10,000 ft away? Because one takes a great deal of contextual information into account. One compares the sign to the trees in its vicinity; and one knows, on independent grounds, that trees (at least outside of northern California) are usually between 15 and 80 ft tall: one can thus use these trees as benchmarks. And one uses various other benchmarks (e.g. houses that one sees, whose sizes and locations one knows one independent grounds) and, more generally, various other pieces of background or contextual knowledge.

By exactly analogous arguments, it can be shown that *any* episode of taking in information about the external world – any episode of taking in the height or speed or shape…or color of an object – involves the co-operation of contextual or background information. Such ulterior knowledge is needed to give content to one's sensory episodes. Otherwise those episodes are indeterminate. If the contextual information that we just spoke of is left out, one sensory episodes – one's visual sensations – don't tell one whether one is seeing a thousand foot object ten-thousand feet a way or a ten foot object a hundred feet away; they don't tell one whether he is seeing a diamond or a square; an oval or a circle; a nearby object moving 3 mph or a distant one moving at 150,000 mph. In the absence of background empirical knowledge, our sensory episodes become indeterminate to the point of meaninglessness – to the point of having *no* representational content.

It seems to me that this Sellarsian line of thought was put forth by Berkeley in 1709. We think we can simply *see* whether an object is ten feet away or a thousand feet away, whether it is oval or round. Of course, we *do* learn objects' shapes, sizes, and so on, through perception, and through sight

in particular. Berkeley doesn't deny this for a second.[260] But he says that when we take in the shape of an object through sight, that is only because a great deal of background knowledge is co-operating.

Berkeley begins by asking to us consider distance-perception. We think we *see* how far away objects are. Obviously, we can (up to a point) take in objects' distances from us through sight. But one obviously cannot tell, just by looking at it, that the moon is this or that distance away (section 44). This is always true, in a very obvious way, of very distant objects. Given *only* one's visual experience, one would have no idea whether the moon was a million light years away and very large, or a few miles away and very small.

Berkeley then says (correctly, in my view) that what is true of very distant objects is true, though less obviously, of not so distant objects. Given *only* what your eyes tell you – leaving aside background knowledge – do you really know that you are seeing a medium sized object fifty feet away or a massive object a mile away? What if you took a guess? What if you guessed that it was fifty feet away, but it took you half and hour to jog to that object? No matter how good your eyes are, and no matter how 'standard' are the conditions of perception, there is no reason why things *couldn't* turn out that way. Your visual experience *by itself* is ambiguous between your seeing a massive object very far away and a medium sized object nearby. So you don't really *see* how far away an object is. When you *think* you are doing that, you are really reading into your visual experience a lot of background empirical knowledge.

You know, for example, that objects of *that* particular shape (say, a stop-sign shape) are typically of such and such a height, and you use *that* information as a kind of benchmark. *Given* that background information, you can tell on the basis of your visual sensation that you are seeing an object that is less than a hundred feet away. But the visual sensation by itself is

ambiguous between a nearby ten-foot tall stop-sign and a far-off hundred-foot stop-sign.

Another example might help. The *appearance* given by of by a receding object is not different from the appearance given off by a shrinking gives. But when you see a person walk into the distance, you can *see* that he is not shrinking. How can you *see* this, given that the appearance in question could be that of a shrinking person? You know that people don't shrink. You also know that if, by some miracle, a person *were* suddenly shrinking, he wouldn't make the calm ambulatory motions made by the person you are seeing. Thus contextual information disambiguates what would otherwise be an ambiguous visual experience.

According to Berkeley, what *ultimately* endows one's visual sensations with information regarding the distances of objects is that different kinds of visual sensations have different kinds of kinesthetic and tactile concomitants. One reads these into one's current visual sensations and it is *this*, in the final analysis, that gives your visual experiences the representational contents that they have. It is *this* that makes your current visual experience be one of (say) a man who is walking away as opposed to a man who is shrinking.

A different example might help. You see what you take to be a pencil lying ten feet away from you. But you find that it takes you half an hour to walk there. And once you are there, that 'pencil' gives off the appearance of a fallen (pencil-shaped) tree. Finally, you return back to your starting point. Obviously you will then 'recalibrate'; you will use the knowledge that you just acquired as a way of establishing the comparative distances of other objects. It must be pointed out that you acquired knowledge of the distance of the pencil *not* merely through sight, but with the help of kinesthetic and tactile sensations, and also sensations of duration. As we just said, you use your knowledge of the distance (from your original starting point) of the

pencil as a kind of reference point: that information helps you know the comparative distances of other objects you are seeing. So, ultimately, it is not the visual sensation *by itself* that apprises you of the distances of all these objects; it is your visual sensation *plus* various sensations of motion, resistance, touch (your feet touching the ground), and duration. If it weren't for these non-visual concomitants of your visual perceptions, you would have no idea how far objects were from you: for your visual sensations *by themselves* cannot give you this knowledge.

Berkeley's basic point is that your visual experience *by itself* doesn't apprise you of the size or distance of the object. What does that is your visual experience *plus* various pieces of background knowledge:

"[T] judgment we make of the distance of an object viewed with both eyes is entirely the result of experience. If we had not constantly found that certain sensations [accompanied our visual sensations], we would never make those sudden judgments from them concerning the distance of objects; no more than we would pretend to judge of a man's thoughts by his pronouncing words we had never hear before." (section 20)

Berkeley sometimes expresses this by saying that you simply don't *see* distances at all (section 46). This seems an overstatement. In *some* sense, surely, I can *see* that my pencil is less than 90 ft away. But there is a deep truth in these words of Berkeley's. One's visual experience, taken by itself, stripped of any contextualization, doesn't tell one that is one is seeing a tree-size pencil a half-mile off or a tiny pencil nearby.

Berkeley generalizes this point (sections 121-147). Consider the kind of appearance that a square object gives off when looked at from directly above – a paradigmatic-'square' appearance (not the diamond-appearance that such

an object gives off when looked at from oblique angles). Accordingly to Berkeley, it is only 'past experience' that teaches us that *that* kind of appearance betokens a square object. Berkeley's argument is cogent. Let S be the kind of appearance just mentioned. (So S is the kind of appearance given off by a square-shaped object looked at from above, under optimal lighting conditions.) Consider the kinds of tactile and kinesthetic sensations that we *in fact* have in connection with a *diamond*-shaped object -- the kinds of sensations one has, for example, when running the palms of one's hand along the periphery of such an object. Let us refer to those sensations as D. And suppose that you had *D* in connection with S. In that case, Berkeley says, S would betoken a *diamond*-shaped object, and not a square one.

(Berkeley thus has a very definite answer to Molyneux's question: Consider a man born blind who was one day given sight. Would he know on the basis of his visual perceptions – before having a chance to correlate them with tactile or kinesthetic or other sensations – whether he was seeing a round or an oval or a square object. Berkeley explicitly says 'no'. (Section 132.))

Berkeley gives analogues of this argument for *every* perceptible property an object might have. For example, it is not one's visual perceptions *by themselves* that tell one whether one is seeing a slow or fast moving object. If one's visual sensations are to apprise one of how fast some object is moving, they need the assistance of a great deal of ulterior information. (If I thought that the plane I saw was only ten feet above me, then I would think it was only moving 3 mph. Since I know that it is thousands of feet above me, I can 'see' that it is moving very quickly.) Berkeley doesn't explicitly make this point in connection with our perceptions of mass or of temperature. But it would be easy enough to adapt what he says about distance, shape, and so on, to apply to these other kinds of properties.

Berkeley illustrates his major point with some clever analogies. When you

hear somebody talk, you feel as though you *directly* grasp their meaning. Somebody says “it’s a lovely day”. You hear *through* the sounds to the meaning. So it is as though your auditory experience *by itself* apprised you of the significance of the sounds you were hearing. But of course that is not so. The relationship between the *sounds* ‘it’s a nice day’ and the corresponding meaning is arbitrary, and you could only know it through experience (you’ve been brought up in an English speaking country, or taken English classes). You ‘hear’ the meaning only with the help of a great deal of background knowledge. You are indeed aware that the person is asserting the proposition *it’s a lovely day*; but it is obviously not your auditory experience *by itself* that does this – it is your auditory experience plus a lot of background knowledge.

Berkeley says that your visual experiences apprise you of the squareness of objects only in the sense in which your auditory perceptions apprise you of the meaning of a spoken sentence. Once again, let S be the kind of visual experience that a square object gives off, when look at from (a certain distance) above, under good conditions. And suppose that O gives of S. You think you just *see* the shape of O – that your visual experience *by itself* tells you what shape O has. But you are reading a great deal of background information into your visual experience: at some cognitive level, you are indeed aware that O is a square, but not *merely* by virtue of the fact that O gives off S. Also at work is your knowledge of the tactile and kinesthetic concomitants of experiences of S.

Berkeley gives another analogy. When a person’s face reddens (in a particular way) that indicates shame. When you see that sort of reddening, you feel as though you can actually *see* the shame. But surely your visual experience by itself doesn’t apprise you of the other person’s shame; a great deal of background knowledge – of how people’s emotions are given bodily expression – is at work. Similarly, when a visual experience makes you

aware of the distance of some object, it is by way of contextual or background information: the experience *by itself* is ambiguous between a 10,000 foot stop-sign that is miles away and an eight foot stop sign that is 20 ft away – and that visual experience could indicate any number of other possibility. On the basis of cultural and other background information (you know that stop-signs are in fact 8 ft tall) the visual experience is 'disambiguated'. Also, you have established, on independent grounds, that some neighboring tree is only ten feet tall, and you use that information as a kind of benchmark. The point is that it is the visual experience *plus* ulterior empirical information that apprises you of the height of the tree – just as it is visual experience of the person's facial reddening *plus* ulterior knowledge of psychosomatic correspondences that apprises you of the person's emotional state.

Berkeley's major point is that no sensory experience *by itself* tells one anything. Any such experience has to be put in the context of a lot of background knowledge *if* it is to bear any determinate message and *if*, therefore, it is to tell one anything about the external world. So no sensory experience is foundational: the senses tell us about the world only through the 'mediation of various past experiences' (an expression Berkeley himself uses frequently). So there is no *given*; there are no *foundational* sensory experiences. Knowledge of the external world is obtained by putting one's current sensory experiences into the context of other sensory experiences and also in the context of previously obtained empirical knowledge. So empirical knowledge is obtained through *coherencies* obtaining among sensory experiences; thus it isn't given through isolated sensory experiences; and thus, for Berkeley, empirical knowledge has no foundation, to use Sellars' own expression. At bottom, this is just what Sellars was saying: it *is* Sellarsian anti-foundationalism.

Actually, Berkeley's argument for what I am alleging to be epistemic anti-foundationalism is far more lucid and developed than Sellars' version. There are, if I am not mistaken, two respects in which Berkeley's argument is superior to Sellars'. First of all, Sellars only gives one example of how sensory experiences need the help of background knowledge to tell us about the external world: one example relating to perception of just *one* kind of a property – and a secondary property (color) at that -- under *non-standard* conditions. Thus, if Sellars' argument for anti-foundationalism is to be made cogent, it must be generalized to *other* properties – e.g. distance, shape – and shown to apply to *standard* (or, at any rate, not grossly unusual) conditions of perception. (In this paper, I tried to provide such generalizations; but, strictly speaking, I went beyond anything that Sellars himself said.) But Berkeley's argument is not subject to these limitations of scope; his argument *directly* shows some kind of anti-foundationalism to hold for our perceptual intake of virtually *every* property (at least, every perceptible property: obviously properties like being composed of neutrinos are excluded) under *standard* conditions. So Berkeley's argument doesn't require much supplementation to be general. (In a *New Theory of Vision* Berkeley doesn't talk about secondary properties.)

There is another respect in which Berkeley's argument is more developed than Sellars'. Sellars argues (rightly, in my view) that to ascertain the color of the tie you need existing empirical knowledge – you need to know something about the conditions of perception. But of course, the question arises: how do you acquire this *existing* empirical knowledge; a kind of regress appears to threaten. Sellars (164—170) gives a very condensed and obscure answer to this. I cannot assess the merits of what Sellars says here. But Sellars doesn't really explain how *representational* mental entities (perceptions) are forged out of what he has compellingly argued to be *non*-representational mental

entities (‘raw feels’, ‘qualia’, ‘sensations’ – all of these are terms Sellars himself uses). Sellars tries to stop the regress *not* by showing how ‘raw feels’ are synthesized into perceptions, but in some much more obscure way, which (frankly) I don’t fully understand. But Berkeley goes to great lengths to show how non-representational mental entities (qualia) are integrated into representational ones; Berkeley talks, in a detailed way, about how a confluence of non-representational mental entities belonging to different sensory modalities can bring about a representational entity (a bit like the way that looking through *both* lenses of a binocular microscope gives you a three-dimensional picture by combining two two-dimensional pictures). Obviously, I cannot off the cuff assess the merits of this aspect of Berkeley’s argument. But Berkeley has provided much *more* of a solution than Sellars’ – I myself am open on whether it is a correct solution -- to the regress that anti-foundationalist seems to incur. Actually, where many contemporary anti-foundationalists rely on the paper currency of obscure metaphors, Berkeley gives us the gold-backing of non-metaphor. For example, knowledge is often said to be an “arc”, in which the various parts support one another, not a sky-scraper, in which some parts support others without being supported by them’(Blackburn 1984:161; Bonjour 1985). Our knowledge is said have the structure of a “four-dimensional curve”; and that, we are told, is why anti-foundationalism isn’t viciously circular or viciously regressive (Bonjour 1985). These images may contain deep truths. But, as they stand, they are just metaphors. Berkeley has provided a very non-metaphorical way of interpreting these metaphors.

It is important in this connection to distinguish two problems that arise on Berkeley-Sellars anti-foundationalism. One is the question how, on any anti-foundationalist view, any belief manages to be *justified.* The other is how anything *representational* (perceptions) manages to arise out of what is not

(so Berkeley and Sellars argue) representational, namely *qualia*, ‘raw feels’, ‘sensations’. It seems to me that Sellars, in the aforementioned obscure pages (164-170) deals to some extent with the former problem, but not with the latter problem: and the latter problem is the more fundamental one. For until we have perceptions, the problem of justification doesn’t even arise. Berkeley doesn’t deal with the problem of justification, but he *does* deal with the problem of how perceptions arise out of non-representational sensations. For reasons discussed earlier, I think that his solution must surely contain some truth, though it appears to be incomplete.

Conception as Knowledge of Uniquely Individuating Descriptions

I.

Let me start by stating the picture I wish to call into question: a picture that is now an orthodoxy, thanks to some of the brilliant arguments found in Kripke's *Naming and Necessity*.

Many a person can refer to Socrates but not be able to produce a description that applies to Socrates and not also to many others. One can use the word "Socrates" with understanding and *not* know the truth of any proposition of the form: *somebody uniquely has phi and "Socrates" refers to x.* When asked who "Socrates" refers to, they may say "he was a great philosopher of ancient times" or "he wrote the Nichomachean Ethics"; and that may exhaust what they have to say (truly or falsely) by way of an answer to your question. What they "know" may be excessively general ("he was a great philosopher of ancient times") or it may simply be false ("he wrote the Nichomachean Ethics"). In both cases, they will lack knowledge of any truth claim of the form: *somebody x uniquely has phi and "Socrates" refers to x.* They will lack knowledge of the right kind of *uniquely individuating description.*

But such people may still be able to produce sentences of the form "… Socrates…", and really *mean* them. So apparently one can refer to Socrates, and also think about him, without knowing of any description that singles him out.

The purpose of this essay is to argue for two related (but probably not coincident) points. First, in order to *refer to* an object, one *does* have to know of some description that singles it out: one has to have knowledge of the right kind of *uniquely individuating description*. Second, in order to *think about* an object, one *does* have to know of some description that singles it out. Without knowledge of a *uniquely individuating description*, there is no reference and there is no conception.

In many cases, the ability to refer to, or think about, an object *does* involve one's being causally connected to the object. But the causal connection, though perhaps essential, nonetheless has a subordinate role: it facilitates conception and reference *by way of* being an indispensable part of one's knowledge of a uniquely individuating description.

I have no doubt that Kripke's major point about proper names is quite correct. "Socrates" is not a disguised definite description. Further, I concede that *part* of what makes tokens of "Socrates" refer to Socrates is (or, more accurately, *can* be) that those tokens stand in a certain causal relation to Socrates himself. Finally, I concede that part of what enables one to think about Socrates is (or, more accurately, *can* be) that one stands in a certain causal relation to the man. But I will argue that these causal relations are always embedded *in* knowledge of a description that applies to Socrates and Socrates alone –knowledge of a *uniquely identifying description.*

None of Kripke's important semantic insights requires an abandonment of the position that, if one is to think about x, one must know of some description or concept that applies uniquely to x. Indeed, I think that when we delineate what Kripke says, we will find that, on the contrary, Kripke's story *requires* that one always have such knowledge.

§ I should make it clear what I will and will not be attacking, and it will

help if I state my biases. First, I believe that direct reference theory is correct – that the sole semantic value of "Socrates" is Socrates. I do not think that, semantically, "Socrates" is a description. I do not advocate any kind of descriptivism with regard to the expression "Socrates". (I will argue that it is not *necessary* to do so in order to accommodate the intuitions that motivate descriptivism.) So, in so far as it is concerned with *semantic* issues, this essay will operate very much within a Kripkean – or, at any rate, a direct-reference – framework.

But semantics is not epistemology. Given only that "Socrates" is a label, and not a description, it does not necessarily follow that I can *think* about Socrates without knowing a description that singles him out. And, I will argue, thinking about Socrates *does* require knowledge of such a description. Indeed, being able to *refer* to Socrates requires knowledge of such a description – notwithstanding that "Socrates" itself has little[261] or no descriptive content at all. The *semantic* relationship between "Socrates" and Socrates is "pure" and non-descriptive. The relation between my concept of Socrates and Socrates is a different story.

A consequence of these findings is that a popular and seemingly plausible analysis of conception is quite false. In fact, much contemporary philosophy of mind and epistemology is a projection of views that *seem* to have support in *Naming and Necessity*, but in fact have none.

§ Let us start with a very quick – and, inevitably, unsatisfactory -- summary of some of the relevant points in *Naming and Necessity*. According to Russell, "Socrates" is synonymous with a definite description; this might be "the greatest philosopher to have died of hemlock poisoning". To *understand* the word "Socrates", and thus to understand sentences of the form "…Socrates…", one must obviously know the semantics of "Socrates". So

one must know that “Socrates” is synonymous with that definite description. Consequently, anyone who uses the word “Socrates” with understanding is able to produce a description that applies to Socrates and Socrates alone. In general, if one knows the semantics of a referring term, then one knows some description that applies uniquely to the referent of that term. If I can use the word “Gilgamesh” with understanding, that means I know of some description that singles out the individual in question. If I know of no such description, then I cannot really know what is meant by sentences like “Gilgamesh was wise”.

Supposedly, Kripke demolished this view. Consider *any* description of Socrates that applies to him uniquely, i.e. that truly singles him out. Here is an example: “the greatest philosopher to have died of hemlock-poisoning”. One can understand sentences of the form “…Socrates…” without knowing that the referent of “Socrates” consumed hemlock. One can understand “Socrates was wise” without knowing that “Socrates” picks out somebody who uniquely satisfies the description: *greatest philosopher to have died of hemlock-poisoning.* This point holds *mutatis mutandis* for any description that singles out Socrates. Consider the description “the person who figures as the protagonist in most of the platonic dialogues”. One can understand the sentence “Socrates was wise” without having the slightest idea that he was a protagonist in any philosophical dialogues. Given *any* description D that applies uniquely to Socrates, one can understand “…Socrates…” without knowing that D applies to the referent of “Socrates”.

This point has many apparent consequences. Some of these are semantic in nature. Others are epistemological.

Here is one of the semantic consequences. “Socrates” is *not* synonymous with any definite description. Given *any* description D, one can know that “Socrates was wise” is true without knowing that somebody uniquely

instantiated D and that any such person was wise. So it isn't tenable to maintain that "Socrates" is synonymous with "the greatest philosopher to have died of hemlock-poisoning" or with any other definite description. (As I said earlier, I accept this.)

Here is one of the supposed *epistemological* consequences. One can *think* about Socrates without knowing of any description that singles him out – without knowing a UID. (As I said earlier, I will argue *against* this.)

§ There are other apparent consequences, both for semantics and epistemology. Let us start with a question: Why does "Socrates" refer to Socrates and not, say, Plato? Russell's answer was this: "Socrates" is synonymous with a description that is true of Socrates and Socrates alone. Kripke made a powerful case against this answer. So how does *Kripke* answer our question? His answer is basically this: "Socrates" refers to Socrates because a certain kind of *causal* relation holds between tokens of "Socrates" and Socrates himself (or, more exactly, states of affairs involving Socrates).

Kripke has more to say on this matter. Somebody – say Smith -- *saw* Socrates. So Socrates' presence *caused* Smith to have certain visual experiences. Smith then gave a name to Socrates. Smith was able to pass that name along to people who did not themselves see, or otherwise sense-perceive, Socrates. Of course, for Smith to pass along that name to someone, that person had to be causally connected to Smith. The process just described is repeated over and over. I myself am the 521st installment in a sequence of events like the one described. Inevitably, Smith's original term for Socrates has undergone phonetic distortion. What I have been left with is the sound (or sound-type) "Socrates". When I produce instances of that sound-type, my words refer back to Socrates precisely because I am connected to Socrates by

way of a chain like the one just described. When I say “Socrates”, my words refer to Socrates because they are causally connected to him in a certain way. In general, if an expression E refers to an object O, that is because tokens of E stand in a certain causal relation to states of affairs involving O.

II.

Let us now evaluate the various claims just put forth. How do you learn what a given word refers to? There appear to be two ways: you can learn it by description and by ostension. (Soon we will see that there is a third.)

You are young; so you have never heard the name “Socrates”. Somebody says to you:

(*) “ ‘Socrates’ is the name of the most famous philosopher to die of hemlock poisoning.”

Here the word “Socrates” has been defined for you in a *descriptive* manner. In effect, you were told: somebody O uniquely satisfies a certain description -- *most famous philosopher to die of hemlock poisoning* – and “Socrates” refers to, or names, O.

On the face of it, there seems to be a second, entirely distinct way of learning what a term refers to: you can be given an *ostensive* definition.

You are in a crowded market place. You see a man haranguing a crowd. Your companion says:

(**) “*that* person over there [pointing to the orator] is named ‘Socrates’”.

My own view is that ostensive definition actually collapses into a kind of descriptive definition. I will argue for this later on. But, of course, this is not a point that can be taken for granted. Right now, we need only register the fact that one can learn what a term refers to through descriptive and also through ostensive definition.

When you give a descriptive definition of a referring term, the descriptive information must be taken to have *wide-scope*. Otherwise the definition turns out false, and Kripke's compelling modal points are violated. (*) defines "Socrates" by means of a description: *most famous philosopher of antiquity to die of hemlock poisoning*. But the descriptive information in (*) is used to *fix the referent* of "Socrates"; that descriptive information is not itself part of the *meaning* of "Socrates".

The following statement is true:

(s1) "Somebody O was a unique most famous philosopher of antiquity to die of hemlock poisoning and, for any predicate *…x…*[262], *…Socrates…* means:…*O*…

The following statement is false:

(s2) "…Socrates…" means: *Somebody O was a unique most famous philosopher of antiquity to die of hemlock poisoning and…O…*

By way of clarification: "Socrates smoked" does *not* mean: *somebody O was a unique most famous philosopher of antiquity to die of hemlock*

poisoning and O smoked.

But there is some object O such that O was a great philosopher who died of hemlock poisoning and such that "Socrates smoked" means: *O smoked.* "Socrates smokes" means *Socrates smoked*. And Socrates was a great philosopher who died of hemlock poisoning. So there is an object O such that O was a great philosopher who died of hemlock poisoning and such that "Socrates smoked" means: *O smoked.*

The descriptive information in (*), and in (s1), is given wide-scope with respect to the "means that" operator. That is why such statements are true. And that is also, of course, why *Socrates drank hemlock* is not analytic and is also contingent (false in some, but not all, possible worlds).

§ It *seems* that, in a certain class of cases, one is able to acquire the ability to use a referring term *without* ever having been given either an ostensive or descriptive definition of it. Suppose you and two other people (Smith and Jones) are talking. All of a sudden Smith and Jones start talking about somebody named "Argo". They never say who "Argo" refers to. They never say: "Argo was the president of our fraternity back in college" or "Argo was my divorce lawyer; and by sheer coincidence he's Smith's racquetball partner." They do not *ostensively* define "Argo"; for he is not present. Smith and Jones simply prattle on about "Argo"; and you are too polite to ask them who that person is.

Nonetheless, as Gareth Evans pointed out, it seems that you *can* still use the word "Argo" in a meaningful way. You can say "it sure sounds like Argo is an awful person"; and you can ask questions like "did Argo mean to do that to you, or was it just an accident?"

So *prima face* it seems that you can acquire the ability to use "Argo" in a meaningful way *without* ever having been given an ostensive or descriptive

definition of it. When you say “it sounds as though Argo is an evil person”, you are using the expression “Argo” with understanding; you are not in the same category as a parrot that mindlessly says “Argo was a cad”. So it seems as though you can simply “pick up” the ability to use “Argo”, and names generally: there needn’t be an ostensive or descriptive definition. [263]

§ Let us take stock. There seem to be three ways that a proper name N can be inducted into one’s lexicon:

(i) one can be given an ostensive definition of N;
(ii) one can be given a descriptive definition of N; or
(iii) one can just “pick up” the use of N.

§ Now let us further consider each these three categories. Suppose that somebody *ostensively* defines “Socrates” for me. I see somebody haranguing a mob. My companion says: “*that* person is named ‘Socrates’”.

Under this circumstance, I know that there is some individual x such that x uniquely has certain properties (he is uniquely a person that I am seeing on this particular occasion behaving in a certain way, dressed in certain clothes, and so on), and I know that “Socrates” refers to x.

Ostensive definitions work by way of knowledge of uniquely individuating descriptions. Somebody *points* to some person (or object) O and says: “*that* guy is named ‘Fred’”. For the definition to work – for you to learn what “Fred” refers to – you must have some way of singling out the person ostended. Suppose your companion points in a certain direction and says “that is Fred”, but there are thirty different equally salient objects. In that case, the definition won’t work. You must be able to pick out the right object.

Picking out a single object involves knowing a uniquely individuating description. An extreme case will bring out my meaning. Suppose that there are thirty numerically different, but (leaving aside location) qualitatively identical men in a certain place. Your companion Bob goes up to one of them, put his hand on his shoulder, and says "this guy is named 'Fred'". Even though the person just named is qualitatively just like many others, you know of some description that applies to *him* and to no one else. You know that: There is some individual x such that x is uniquely a person whom Bob just picked out – such that x uniquely has the property that Bob just put his hand on his shoulder – and "Fred" names x.

Now imagine a slightly different scenario. You and Bob are in the presence of thirty qualitatively identical individuals. (They are in different places of course.) Bob doesn't actually put his hand on anyone's shoulder. But there is some person x such that Bob *is* able to use an act of pointing to distinguish x from the other people (and objects) in his vicinity and such that Bob tells you that "Fred" refers to x.

Bob's definition works. It works only because you know of some existence claim that Fred, and Fred alone, satisfies. You know that: there is some specific person x such that x uniquely is such that Bob just pointed him out. You also know that "Fred" refers to that person. But your being able to know "Fred" refers to that specific person involves your knowing the truth of some claim of the form: *somebody x uniquely has phi and "Fred" refers to phi.* You know what "Fred" refers to only because you know of some description that singles out Fred: some "uniquely individuating description".

The description is not one that tells you much about what kind of person Fred is. In fact, that description doesn't tell you *anything* about how Fred is psychologically or (leaving aside facts about spatio-temporal location) physically different from other people. The description in question gives you

enough information to cognitively single out Fred, and it gives you very little, if anything, beyond that. But that is neither here nor there. In this context, there is only one important point. Your knowing what "Fred" refers to depends essentially on your knowing the truth of some proposition of the form: *somebody x uniquely has phi and "Fred" names x.*

This is not to deny that your knowing what "Fred" refers to involves your having some kind of causal connection to Fred. You *do* have some kind of causal connection to Fred: you *see* him, and this involves your being causally connected to him. You acquire the ability to use the word "Fred" in consequence of (among other things, I will argue) this causal connection. So your subsequent uses of "Fred" refer to Fred in virtue of, possibly among other things, this causal connection.

But the causal connection is not the primary thing; the causal connection is a component of something more important. It is a component of your knowledge of a uniquely individuating description. How is the meaning of "Fred" given to you? It is given to you ostensively. But, even so, it is, at least in effect, given to you through an existence claim. You know the truth of some claim of the form: there is some x such that x uniquely has phi and "Fred" refers to x. Now, your knowing the truth of that claim involves your being causally connected to Fred. For the claim in question involved *visual* perception (or, at any rate, some kind of sense-perception); and seeing (or otherwise sense-perceiving) a thing involves being causally connected to it. But *how* is vision (sense-perception) involved? It is involved in exactly this sense: it enables you to know the truth of some existence claim of the kind described a moment ago; sense-perception is involved in the sense it apprises you of the truth of a *uniquely individuating description.* A causal connection to Fred is involved in your knowing what "Fred" refers to: but this causal connection is involved only to the extent that is constitutive of your knowing

the right kind of uniquely individuating description.

§ What we've established so far is this. In *some* cases, not all, we learn what a term refers to through ostensive definition. When we acquire the ability to use a term through ostensive definition, our ability to use that term with understanding consists in our having knowledge of some uniquely individuating description of the referent: it involves our knowing the truth of some claim of the form *something x uniquely has phi and N refers to x,* where N is the word in question.

§ Here one point should be made explicit. One's uniquely individuating description of an object need not be, and almost never is, "pure": there is almost always an indexical component. Suppose Bob points to somebody and says "that is Fred". Here, as we've just seen, I assign a referent to "Fred" by way of my knowledge of some claim of the form: *something x uniquely has phi and "Fred" names x.* In this case phi would be some property like: *man wearing plaid shirt over there, on whom I am uniquely focused.*…The uniquely individuating description will thus be loaded with context-sensitive components (whose linguistic counterparts are indexicals). This fact might lead to a conclusion. Suppose my doppelganger on twin-Earth is in a situation exactly like mine except that he is talking with twin-Bob and he is seeing twin-Fred (who is standing next to twin-tree, and so on). It might seem that the uniquely individuating description through which I acquire a concept of Fred is identical with that through which my twin acquires a concept of twin-Fred: both descriptions are linguistically expressed by the same phrase: "the unique person wearing the plaid shirt…on whom I am at this moment focusing…" Thus, it might appear that neither my conception, nor my twin's, is *uniquely* individuating. A Strawsonian "massive reduplication" has

occurred.

So it seems that my uniquely supposed individuating description applies equally to Fred and twin-Fred – and thus isn't a *uniquely* individuating description at all. The same is true *mutatis mutandis* of my twin's "uniquely" individuating description of twin-Fred. Thus it seems that a uniquely individuating description is *not* enough for conception. The reason I acquire a concept of Fred, and not twin-Fred, is that I am *causally* connected to Fred. So, once again, the causal theory of conception trumps the *descriptive* theory of conception.

This viewpoint is based on a simple failure to understand some points about demonstratives that were clearly articulated by Frege, and improved by Kaplan, Perry and others. On Monday, in Amsterdam, I say "the man over there is rich". Focus on my token of "the man over there". On Tuesday, in Madrid, I say "the man over there is rich". I have used the expression "the man over there" twice. But different uniquely individuating descriptions were expressed. In the one case, I was talking about the man in place p, at time t (where p is some part of Amsterdam, and t is some time on Monday). In the other case, I was talking about the man in place p*, at time t* (where p* is some part of Madrid, and t* is some time on Tuesday). Different uniquely individuating descriptions were expressed and were thought. *Mutatis mutandis* this point applies exactly to the scenario described concerning me and my twin.

Now it *is* true, at least in part, that I upload a concept of Fred, as opposed to twin-Fred, because I have a *causal* connection to Fred that I don't have to twin-Fred. But notice that this causal connection enters into the story *by way of some feature of the relevant uniquely individuating description.* That description has an indexical component; and the causal connection fixes what that component picks out. So the causal connection is indeed essential: but it

is essential *as* a mere component of a fundamentally descriptive epistemic modality.

III.

As we saw, there seem to be three ways that a proper name N can be inducted into one's lexicon:

(i) one can be given an ostensive definition;
(ii) one can be given a descriptive definition of N; or
(iii) one can just "pick up" the use of N.

We've already dealt with (i). Let us now deal with (iii). We will soon see why we are skipping to (iii), and not dealing right away with (ii).

Once again, consider the case where you are talking with Smith and Jones. They suddenly starting talking about somebody named "Argo". They are acquainted with Argo, but you are not. At no point in the conversation do they ever say anything like "Argo was my divorce lawyer" or "Argo was the best man at my wedding". They never enunciate a uniquely individuating description; they never give you a *definite* description that applies to Argo.

Of course, given that they are producing sentences of the form "… Argo…", it follows that you are, in effect, given many *in*definite descriptions that apply to him. Smith says "Argo is a real cad." So here you are being given an existence claim of the form: somebody x is a cad and "Argo" refers to x. But, of course, many people are cads; so given only that you know the truth of this existence claim, it doesn't follow that you know which *specific* person – which *specific* cad – "Argo" refers to.

But the truth is: merely by virtue of the fact that you overheard Smith and Jones using the word “Argo”, you *have* been given a *definite* description that applies to Argo, and you *have* learned some true proposition of the form: some thing or some person x uniquely has phi and “Argo” refers to x.

Unlike you, Smith *is* personally acquainted with Argo. Thus, for reasons we’ve covered, Smith knows some true proposition of the form: some thing (or some person) x uniquely has phi and “Argo” refers to x.

In the course of your conversation with Smith and Jones, you often hear Smith make statements like “Argo was a cad”, and “maybe Argo wasn’t so bad”. On each such occasion, there was *one* person to whom Smith was referring by means of the word “Argo”. Suppose that, on occasion *t*, Smith says “Argo was a cad”. Obviously the description “x is a cad” is indefinite, and applies to many. But on that occasion, there was exactly one person whom Smith was describing as being a cad; and you know this, merely by virtue of overhearing Smith’s statement. So, just by virtue of being in colloquy with Smith, you *do* pick up knowledge of some claim like: on occasion *t*, there was exactly one person who Smith described as being a cad, and “Argo” refers to x.

§ To test the mettle of my thesis, let us make things as difficult as possible for it. Suppose that Smith goes on a rant. He says “Flipper is a cad; Argo is a cad; Ichabod is a cad; Peewee is a cad…” You have never heard *any* of these names before; you have not even heard Smith (or Jones) use any of those names in the context of the conversation in which you are currently engaged. Let us further suppose that you never hear anyone use any of these names again; so on no later occasion are you given any further information about them or their referents. Nonetheless, merely by virtue of overhearing these words on that one occasion, you can appropriate them into your lexicon, and

use them meaningfully: you can make statements, and ask question, *about* Argo and Flipper and Peewee; for example, you can say “Peewee and Argo probably didn’t have very nurturing parents”.

This fact *seems* to confute the theory I just put forth. You hear Smith use the name “Argo” just once; and what Smith is saying about “Argo” is identical with what he is saying about “Peewee” and “Flipper” and all the rest. So, under these circumstances, you do *not* know some true existence claim of the form: *on such and such occasion, there was exactly one person whom Smith was describing as a cad, and “Argo” refers to x.* For that claim isn’t true; on the occasion in question, Smith was describing several people as cads.

Despite all of this, I submit, you *still* know of some true existence claim of the form. In fact, you know several such existence claims. First of all, we must take care not to define the term “occasion” too broadly. The term “occasion” can refer to a three week period or a three nanosecond period. Suppose you go to three-day long party and you meet many people. You can truly say “I met two hundred people on that occasion”. But *within* that occasion, there were other occasions. Perhaps you went to the veranda and met some people there. On *that* occasion you met forty people. And within *that* occasion, there were yet others. There was the time that so and so spilled his drink: referring to *that* incident, you could truly say “on *that* occasion I seriously regretted leaving my house that day.”

Terms like “occasion” and “context” are (pun intended) context-sensitive. An occasion (or context) could comprise nothing more than a two millisecond, three cubic millimeter of space-time. But an occasion could conceivably comprise an entire galactic era.

Given this, let us reconsider your discussion with Smith and Jones. The statement “on that occasion, there was no *one* person whom Smith described

as a cad" is true, if by "that occasion" you mean the whole discussion. But there were many occasions *within* the discussion. There was the time Jones spilled his beer; the time Smith made an unseemly pass at the waitress; the awkward moment when you realized that you hadn't brought any money with you; the time Smith used the name "Yaroblinsky". Within that discussion Smith described many people as cads. But *given any one of those people*, there was indeed some occasion, within the discussion in question, such that, on that specific occasion, Smith described that person, and *only* that person, as a "cad". Let x be *any* one of the unfortunates who, in that conversation, Smith described as a cad. Within that conversation, there was some occasion such that, on that occasion, Smith described x, and x alone, as a cad. Of course, Smith's referring to x as a cad involved Smith's using some name: "Argo" or "Peewee" or "Ichabod" or "Beezer". Smith had to say "*Argo* is a cad" or "*Beezer* is a cad". He couldn't just say "__ is a cad" or "x is a cad". Let N be *any* one of the names Smith used, in the discussion in question, to ascribe the property of being a cad to somebody; and, as before, let x be *any* of the people to whom Smith ascribed that property. Within the conversation in question, there was some occasion such that, on that occasion, Smith described x, and x alone, as a cad and such that Smith used N to refer to Smith. Thus, within the conversation in question, there was some occasion such that, on that occasion, Smith described some person x, and x alone, as a cad and such that Smith used "Argo" to refer to x. Within that same conversation, there was some *other* occasion such that, on *that* occasion, Smith described some person x, and x alone, as a cad and such that Smith used "Peewee" to refer to x. Exactly the same thing (*mutatis mutandis*) is true of "Ichabod" and "Beezer", and each of the other aforementioned names.

Let us focus on "Argo". You are a part of the conversation in question; in

any case, you overhear it. Within that conversation, there was some specific occasion when Smith described some *single* person x as a cad, and Smith referred to x as “Argo”. Here, of course, the term “occasion” is covering a small window of time. But, uncontroversially, *within* that narrow window, there was some *one* person whom Smith described as a cad, and to whom Smith referred as “Argo”. Since you are part of the conversation, you know this. When you hear Smith say “Argo is a cad”, you know that Smith is referring to a single person; and you can distinguish the occasion – the narrow window of time – in which Smith produced that sentence from the narrow window of time in which Smith said “Peewee is a cad”. And you can distinguish both occasions from the narrow window of time in which Smith said “Beezer is a cad”. You can cognitively keep all of these occasions separate. Given any one of these occasions, you know that Smith was describing exactly one person x as a cad; and you also know which name Smith used to make that statement. So there is some specific occasion O such that, on O, Smith describes a single person x as a cad, and Smith used “Argo” to refer to x. And you know this: you know it merely by virtue of being a part of the colloquy in question. So merely by virtue of being a part of that discussion – nay, merely by virtue of over-hearing it – you know the truth of some claim of the form:

(*) *on such and such occasion, somebody x was uniquely such that Smith described x as a cad, and Smith used “Argo” to refer to x.*

There is only a trivial difference between (*) and:

(**) *on such and such occasion, somebody x was uniquely such that Smith described x as a cad, and “Argo” refers to x.*

So *merely* by virtue of being in a position to "pick up" use of the term "Argo", you are *ipso facto* in a position to know the truth of (**). Thus *merely* by virtue of being in a position to "pick up" use of the term "Argo", you know some existence claim of the form: there is some x such that x uniquely has phi and "Argo" refers to x.

It is plain that what *enables* you to pick up the term "Argo" is your knowledge of some claim of the sort just described. Your picking up a term like "Argo" -- i.e. your appropriating it in manner (iii), as opposed to manner (ii) or (i) – involves your witnessing a sentence of the form "…Argo…" It involves your encountering a case of somebody's ascribing a property to Argo. (This is true even if you read such sentence: you are *still*, in effect, encountering somebody's using that sentence to make such a statement.) Whenever you witness a token of "Argo has phi" you are witnessing an occasion such that, on that occasion, there is somebody x such that, *on that occasion*, the property of being a phi is being ascribed to x and x alone; and you also witness that, on said occasion, "Argo" is being used to refer to x. So if you "pick up" use of the word "Argo", that involves the following: there is some specific occasion O such that you know that, on that occasion, exactly one person x is being described as having phi and "Argo" refers to x.

§ We must be careful to distinguish the following claims:

(a) Exactly one person has phi.
(b) Exactly one person is being *described*, on some particular occasion *t*, as having phi.

There are many cads. But there are many occasions where only one person is being *described* as a cad

Your picking up the word "Argo" doesn't involve your learning the truth of some claim like: exactly one person x is a cad, and "Argo" refers to x. Such a claim is false; there are many cads. Your picking up "Argo" involves your learning the truth of some claim: on some specific occasion, exactly one person x is being *described* as a cad, and "Argo" refers to x.

We must take care to make one other distinction. Smith *himself* never uttered an existence claim of the form: there is some x such that x uniquely has phi and "Argo" refers to x. Smith *himself* never uttered an existence claim of the form: *on occasion t, I (Smith) referred to somebody x as a cad and "Argo" refers to x.* Smith said only "Argo is a cad", "Peewee is a cad", and so forth. But the information that you are given, by virtue of being in discussion with Smith, is not confined to what Smith himself *says.* When Smith *says* "Argo is a cad", the information that is encoded in his *words* is quite threadbare. But you are given information above and beyond that information; you witness Smith's speech act. You put that speech act within a larger context; you have an enormous amount of information about when that speech act occurs. You know that it occurred after certain other events (e.g. before Smith spilled his beer) and before certain other events (e.g. before Jones made the unseemly pass at the waitress). In any case, even though what Smith was *saying* might have been minimal, the information you have *about* Smith's saying it – about the context in which that speech-act occurred -- is enormous; and *that* information does suffice to give you the right kind of uniquely individuating description. Smith's words do not *by themselves* give you a uniquely individuating description that applies to Argo or Peewee or Ichabod. But there is something that *does* give you such a description: that

something is your perception of the conversational context in which Smith uses the word “Argo”.

§ There is no vicious circularity in this account. Smith personally met Argo. He learned the meaning of “Argo” through an ostensive definition. Maybe Smith *himself* gave Argo the name “Argo”. (For cultural reasons, this is not likely. But that is not relevant here.) In any case, Smith *saw* Argo, and then learned (or possibly stipulated) that the thing he was seeing was named “Argo”.

You meet Smith. You have a conversation with him. On some specific occasion within that conversation, Smith says “Argo was a cad”. You thus know that, on that occasion, there is somebody x such that Smith is describing x, and only x, as a cad, and that “Argo” is the term Smith is using to refer to x. the uniquely individuating description you associated with “Argo” is not: “the unique person x such that x is named ‘Argo’”. Rather, it is: the unique person x such that, on such and such occasion, Smith was describing x as a cad.”

In the circumstance just described, you are only at one remove from somebody who was directly acquainted with Argo – with somebody to whom “Argo” was defined ostensively.

But what we’ve said applies equally to the case where you are at five hundred removes from the referent of the word in question. Consider this scenario. Smith doesn’t personally know Argo. But Smith knows Brown, and Brown personally knows Argo. Smith is not given an ostensive or descriptive definition of “Argo”. But Smith overhears Brown saying “Argo is a cad” or whatnot; and on this basis, Smith is able to add “Argo” to his own lexicon: we just described the principles governing such an addition. You then overhear Smith saying things like “Argo is a cad”, “Argo is a mean person”,

or whatnot. You are not given an ostensive or descriptive definition of "Argo". But on the basis of what you overhear, you are able to add "Argo" to your lexicon: the principles that govern *Smith's* appropriation of "Argo" obviously apply to your appropriation of that name. By exactly similar reasoning, it wouldn't matter if you were five hundred removes from somebody who was actually acquainted with Argo: the same principles would apply; and in each case, what would enable you to use "Argo" to refer to Argo would be your knowledge of some claim of the form: something x uniquely has phi and "Argo" refers to x (where phi would be some property of the form: *is a person who, on such and such occasion is being described as having thus and such property*).

§ Of course, words undergo phonetic degradation and alteration as they change hands. The first person who uses the word "Argo" might pronounce it one way; the people who "pick up" that name from that person might pronounce it differently. But this fact obviously doesn't affect the principles governing appropriation of that word. Brown might pronounce "Argo" one way; Smith might pronounce it differently; you might pronounce it a third way; the people to whom *you* transmit this word might pronounce it a fourth way. After a few transmissions, what is left might bear extremely little resemblance to Brown's pronunciation. In some contexts, that fact is of importance. But that fact has no bearing on *our* point: that, whenever somebody "picks up" a name – whenever they add it to their lexicon in manner (ii) – it is in virtue of their having knowledge of a uniquely individuating description that applies to the referent. It is obvious how these points apply to names like "Socrates". The word that Socrates' own acquaintances used to refer to him was not "Socrates"; that word probably had some phonetic similarities to "Socrates"; but there are also non-trivial

differences. Nonetheless, however distorted “Socrates” may be relative to the word that the Ancient Greeks used, the principles governing the “picking up” of that word – or any one of its ancestors or descendants, no matter how much phonetic distortion is involved – are the same. In each case, whenever one adds it to their lexicon in manner (ii) – it is in virtue of their having knowledge of a uniquely individuating description that applies to the referent, namely Socrates.

So suppose that “Sukrat” is the Ancient Greek word for Socrates. Glaucon learns what “Sukrat” means through an ostensive definition. Glaucon is engaged in a discussion with Zenon. Glaucon never gives Zenon any ostensive or descriptive definition of “Sukrat”. But Glaucon *does* produce a sentence of the form “…Sukrat…” Thus, for reasons discussed, Zenon acquires the ability to use “Sukrat” to refer to Socrates; and, for reasons discussed, Zenon’s ability to do this involves his knowing a uniquely individuating description that applies to Socrates. Zenon has a conversation with Phoebus. Zenon slightly mispronounces “Sukrat”. In a manner exactly analogous to that just described, Phoebus acquires the ability to use that, slightly mispronounced version of “Sukrat” to refer to Socrates. In this way, variants of “Sukrat” are transmitted from person to person, from generation to generation. What we are left with, two thousand years later, is a word that is quite different (at least phonetically) from “Socrates”. But one’s acquiring the ability to use that word, however distorted it might be relative to the original word for “Socrates”, involves one’s learning some uniquely individuating description that applies to Socrates; it involves one’s learning some true claim of the form: somebody x uniquely has phi and “Socrates” has x. But, in most cases, “phi” will be some property like *is described, on such and such occasion, by so and so, as having thus and such property*. “Phi” will *not* be a property of historical moment; it will *not* be a property like *great*

philosopher of antiquity who drank hemlock or *philosopher who figured as the protagonist in most of Plato's dialogues.* The property in question will be some highly contextual one, some property that doesn't connect in any significant way with Socrates' intellectual or historical countenance.

IV.

As we've noted, there are three ways you can add a referring term to your lexicon: We've found that whenever (i) or (iii) is the case, your ability to use the word in question involves your having knowledge of a uniquely individuating description of the referent of that word. Let us now consider the second of these three categories: descriptive definition.

Here it is very easy to fall prey to a confusion. Imagine the following. You have never heard the word "Socrates" before. You are talking with your friend Charlie; he says "Seymour is second only to Socrates in philosophical ability". You don't know who "Socrates" refers to? So you ask: "Who is Socrates?" Charlie says: "Socrates was the author of the Nichomachean Ethics and the Posterior Analytics".

Of course, Charlie is wrong; Socrates did *not* write either of those works; and "Socrates" does not refer to the author of either of those works. At the same time, in consequence of your having this exchange with Charlie, you can use the word "Socrates" to refer to Socrates.

Here it *seems* as though you are able to use the word "Socrates" to refer to Socrates, even though your only belief about him is wrong – even though the only thing you "know" about him isn't true. But this appearance is misleading. What we have here is a case where you add a word to your lexicon in manner (iii). You *do* know some uniquely individuating description that applies to Socrates. But that description is not "the unique

author of the Nichomachean Ethics and the Posterior Analytics". Rather, it is: *the person to whom Charlie, on occasion *t*, ascribed the property of writing the Nichomachean Ethics and the Posterior Analytics*[264], for some value of *t*.

In most cases where we *seem* to be appropriating a name in manner (ii), we are *in fact* appropriating it in manner (iii). But we can certainly *construct* a case where, unmistakably, we are appropriating a name in manner (iii).[265] Suppose you and Charlie are having a conversation. Neither of you has any idea who invented the zipper. You and Charlie decide that, at least when speaking to each other, you will refer to that person, whoever it is, as "Julius". So you have no idea who invented the zipper. But you know that somebody did so, and you have stipulated that "Julius" is to refer to that person. Thus you have added a semantic rule to English (or, at any rate, to a certain idiolect of English). It must be made clear what that rule is. The rule is given by the following:

(Z) Somebody x uniquely invented the zipper, and "Julius" refers to x.

The descriptive information is given *wide*-scope with respect to the "refers to" operator. So if Smith was a unique zipper-inventor, then "Julius was a millionaire" means *Smith was a millionaire*. That sentence does *not* mean: *somebody x uniquely invented the zipper and x was a millionaire*. "Julius" is a name, not a quantifier.

You and Charlie don't know who "Julius" refers to (nether do I). But it does refer to somebody. On the face of it, it certainly *seems* as though you can use it significantly: you can, with understanding, say things like "Julius was probably a talented inventor" and "Julius must be worth a fortune".

Some people have said that, under the circumstances, you *cannot* say things like “Julius must be worth a fortune” and really *mean* them. This may or may not be true (I will weigh in on this shortly). But what everybody will agree on is this. *If* you and I can use the term “Julius” with understanding, it is only because we know some true claim of the form: somebody x uniquely has phi, and “Julius” refers to x. Here, of course, phi is the property of inventing the zipper.

§ I do not wish to be misunderstood. A strong case can be made that unless you know *who* invented the zipper, you don’t know which proposition is semantically encoded in “Julius must be worth a fortune”, and other such sentences. What you actually know, it could be argued, is some proposition like: *for some object O*, the sentence “Julius must be worth a fortune” is true exactly if O must be worth a fortune. Since you don’t know which object invented the zipper, you don’t know who “Julius” refers to; and you therefore don’t know which proposition is encoded in that sentence. But, at the same time, you have a kind of metalinguistic knowledge *about* “Julius must be worth a fortune” which, in some respects, simulates real understanding of it.

Be all of this as it may, what is clear is that *if*, under the circumstances described, you or Charlie understand “Julius must be worth a fortune” – so *if*, under those circumstances, you understand “Julius” – it is entirely because you know the truth of some claim of the form: somebody x uniquely has phi [the property of inventing the zipper], and “Julius” refers to x.

Everyday life there are many cases where we *seem* to be adding a name to our lexicon in manner (ii). Suppose little Timmy has never heard the name “Shakespeare” before. You say to him: “Shakespeare is the person who wrote *Romeo and Juliet*”. Here it *seems* that the three year old is appropriating the name “Shakespeare” in manner (ii); for what you are telling him is, in effect:

there is some x such that x uniquely wrote Romeo and Juliet and "Shakespeare" refers to x.

But really little Timmy has appropriated "Shakespeare" in manner (iii). A Kripke-style argument shows this. After your conversation with him, little Timmy can make statements about Shakespeare; they may be wrong and ill-informed; but they are wrong and ill-informed statements *about* Shakespeare. Let us suppose he does so; he asks "did Shakespeare like French fries?", and says "Shakespeare must have been weird, because his writing is so funny".

But what if it turns out that Shakespeare did *not* write *Romeo and Juliet*? What if it turns out that, although he wrote *Hamlet* and *King Lear* and *Timon of Athens*, he didn't write Romeo and Juliet? Does that mean that little Timmy's statements were blanks? Was the three year old not referring to anybody? It seems not; it seems that little Timmy did manage to ask whether *Shakespeare* liked French fries.

But if Little Timmy's ability to refer to Shakespeare depended on Shakespeare's having written *Romeo and Juliet* – if little Timmy's ability to refer to Shakespeare depended on the accuracy of the statement: *somebody uniquely wrote Romeo and Juliet and "Shakespeare" refers to x* – then his statements would indeed have been abortive.

What is going on here is that little Timmy added "Shakespeare" to his lexicon in manner (iii). He heard *you* refer to him; so the existence claim through which he added "Shakespeare" to his lexicon is something like: *there is some x such that, on such and such occasion, Daddy said that x wrote Romeo and Juliet and "Shakespeare" refers to x.* And *that* existence-claim remains true, even though the other one -- *somebody uniquely wrote Romeo and Juliet and "Shakespeare" refers to x* – turns out false.

What role did this other existence claim – the one that turned out to be wrong – have? Suppose you see meet somebody at a party; you don't know

him; but he acts as though he is very important. You ask your companion: "Who is that guy? What gives him the right to be so self-important?" You are told: "He is the dean". Before you know who that person was, you were able to refer to him, and also to have thoughts about him. You were able to think that *he* is obnoxious and to wonder what gives *him* the right to be so haughty. But, despite this fact, there is still *some* sense in which you don't know who he is. You can be able to think about, and refer to, somebody, without knowing quite *who* they are. When you tell little Timmy "Shakespeare is the guy who wrote Romeo and Juliet", you are telling him (or trying to tell him) *who* Shakespeare was: and this, as we've just seen, is different from transmitting to Timmy the ability to think about, or refer to Shakespeare. (Of course, what we've said about Shakespeare and "Shakespeare" applies to inanimate objects and the words that label them.) Most "descriptive definitions" are really cases where one is learning *who* somebody is, or *which* object something is; they are *not* cases where one is acquiring the ability to refer to, or think about, that thing. Usually, the ability to refer to, and think about, the object in question seems to pre-exist knowledge of *who* somebody is or *which* object something is.[266]

There *are* cases where one is genuinely acquiring a name in manner (ii); but they are not easy to produce; and they tend to be somewhat artificial.

But here there is only one really important point. *Whenever* one adds a referring term to one's lexicon, that involves one's knowing some uniquely individuating description of the referent: for any name N, then one's adding N to one's lexicon involves one's knowing some true claim of the form: *somebody x uniquely has phi, and N names x*. (This is true, we found, even in the case where one acquires the name through ostensive definition.) The existence claims usually are not ones of historical moment; they are not ones like *somebody x wrote Romeo and Juliet and "Shakespeare" names x*. They

are usually highly specific to the peculiarities of one's own situation. But what is important is that knowledge of *some* such claim, historically trivial though it may be, is needed to use a name to refer to a thing. I wish to be clear on one point. Even though we learn what a name refers to *through* some existential-descriptive claim, the name itself is just a label. Semantically, "Shakespeare" simply labels Shakespeare; it doesn't have any descriptive content (except such as it has in virtue of being a mere label of Shakespeare). But one always learns *which* thing a name labels *through* some existential-descriptive claim. The semantics are not existential-descriptive; but the act of *learning* the semantics is existential-descriptive. The semantic rule for "Shakespeare" must be *given* to you in some way or other. But what is given is not the same thing as *how* it is given. The following statement is false:

(S1) The semantic rule for "Shakespeare" is this: somebody x uniquely wrote *Hamlet* and "Shakespeare" names x.

S1 is false because "Shakespeare" refers to Shakespeare in worlds where *Hamlet* was never written. So the semantic rule for "Shakespeare" cannot have anything to do with *Hamlet.*

But the *following* statement is true:

(S2) Somebody x uniquely wrote *Hamlet* and, and the semantic rule for "Shakespeare" is this: "Shakespeare" names x.

The operator "the semantic rule for__ is" must be given *narrow*-scope. Otherwise we end up committed to a false descriptivism. The name is just a label. The literal meaning of "Shakespeare was prolific" is simply: *Shakespeare wrote Hamlet.* It is *not*: somebody x uniquely wrote Hamlet and

x was prolific.

Though false, S1 is not, so to speak, *as* false as the following:

(S3) The semantic rule for "Shakespeare is this: for any predicate *…x…*, *…Shakespeare…* means: somebody O uniquely wrote Hamlet and…O…

What is literally meant by "Shakespeare was bald" is simply *Shakespeare was bald*; it is not: *somebody x* uniquely wrote Hamlet and x was bald.

To avoid running afoul of Kripke's correct points, we must take care to give the relevant operators – "refers to", "is the semantic rule for__" – *narrow* scope. We must make sure that they are safely to the *right* of potential descriptive, and therefore modal, contaminants. We've seen that, whenever you learn the semantics of a name, it is *through* descriptive information. But whenever we learn the semantics of a name, we apparently give the descriptive content wide-scope. That, presumably, is why you don't regard *N has phi* as *analytic*, or as true in all possible worlds, even though you learned N's semantics by being told that something x uniquely has phi and N names x.

§ Dummett (1973) argues for wide-scope descriptivism. "Socrates was necessarily bald" means: *there was some x such that x was a unique great philosopher to drink hemlock and necessarily: x was bald.* It is pretty clear, I think, that this doctrine is false. (See Soames 2001 for an excellent discussion.) "Socrates was bald" doesn't entail "somebody drank hemlock". Dummett's analysis deals with some of Kripke's modal points, but not with Kripke's compelling point that there is no *analytic* connection between "Socrates was bald" and "somebody drank hemlock." It will be noticed that (S2) is true only because, in it, the "the semantic rule for__" operator is given

narrow scope, so that no descriptive information falls within its clutches. By being careful about where we put that operator, we avoid any kind of descriptivism --narrow-scope *or* wide-scope --and our analysis remains Kripke-friendly.

V.

There is a point that we must address. It will help if we first define a bit of terminology. Let

(^) something x has phi

be a true proposition. Of course, what makes it true is that, indeed, some *particular* thing O has phi: the truth of any generalization supervenes on the truth of some singular proposition. Let us refer to O as the "verifier" of (^). Consider the claim "somebody is bald". This is true because some *particular* person is bald – because *Fred* is bald or *Bob* is bald. So Bob and Fred are *verifiers* of "somebody is bald". An existence claim can have many verifiers. (False existence claims have none.) A true existence claim of the form "somebody *uniquely* has phi" has exactly one verifier.

Now let us state the problem that my analysis faces. One typically forgets the existence-claim through which one added a name to one's lexicon. One typically forgets the *first* time one heard "Shakespeare" or "Aristotle"; one often forgets one's *first* meeting of a person. Given this, one is tempted to say:

"You don't necessarily associate *any* uniquely individuating description with "Aristotle" or "Smith" or "Jones". Such a description was needed to

enable you to refer to those people, and to think about them. But once that ability is in place, the description drops out. At that point, there can be *pure* thought about the entities in question. You can *just* think about Aristotle and Smith; you don't have to think about them *through* a uniquely individuating description.

This, I believe, is not how it works: conception is not such a psychologically or epistemically impoverished phenomenon – it is replete with descriptive information. What happens, I would argue, is not that one *ceases* to have knowledge of a uniquely individuating description, but rather that one replaces one such description with a related but distinct one.

Remember what we said earlier about "Fred". If you add "Fred" to your lexicon in manner (i), i.e. through an ostensive definition, that involves your knowing some claim of the form: *somebody x uniquely has phi and "Fred" names x.*

Suppose you and a friend are at a party. You see somebody wearing a bowler hat, talking animatedly. At time t, your friend points to that person and says "*that* is Fred". Your adding "Fred" to your lexicon involves your knowing some existence claim, something like:

(e1) *there is now somebody x whom I am sense-perceiving and attending to, and who is wearing a bowler hat, and "Fred" names x.*

Let us suppose that, a minute or so after "Fred" is ostensively defined for you, you go up to him and start talking. Now you have new information about Fred; you have a new uniquely individuating description of him. There are a number of reasons for this. First of all, the mere lapse in time between your current situation – your now talking with Fred – and the initial ostensive

definition by itself guarantees that you will have a new uniquely individuating description of him. At time t, (e1) was your uniquely individuating description. But now that time has passed, (e1) has been replaced. By itself, the mere passage of time replaces (e1) with:

(e2) *there is somebody x whom, a minute or two ago, I was sense-perceiving and attending to, and who was then wearing a bowler hat, and "Fred" names x.*

notice that (e2) "interlocks" with (e1). The thing having the property specified in (e2) is the same as the thing having the property specified in (e1). The uniquely individuating description changes. But the new one is related to the old one in such a way that something x uniquely makes both of them true. The new one "inherits" the verifier of the old one. In a moment we will say more about this process of inheritance or transmission.

But other things, besides the mere passage of time, will replace (e1), with a new uniquely individuating description. You are now engaged in conversation with Fred. You can see him more clearly (although from afar he looked hale and strong, you now see that he is puffy and out of shape); he is saying certain things to you (he is talking about the national debt and the current administration). As a result, you generate a *new* uniquely individuating description of Fred. Your new one is something like:

(e3) *there is somebody x whom, a minute or two ago, I was sense-perceiving and attending to, and who was then wearing a bowler hat; and I am not attending to x, and conversing with him, x is now talking about the national debt, and x looks puffy and unhealthy, and "Fred" names x.*

Notice that (e3) and (e2) interlock; (e2) transmits its verifier to (e3).

Now that you have (e3), you can *drop* (e2) and (e1). You can completely forget them; and that will not jeopardize your ability to think about Fred. Suppose a moment passes; you have stopped talking with Fred. By itself, the mere passage of time replaces (e3) with a slightly different claim:

(e4) there is some x such that, a moment ago, x was talking about the national debt, and x looked puffy and unhealthy, and "Fred" names x.

Suppose that you have completely forgotten (e1)-(e3). You *still* have some uniquely individuating description of Fred. And, so long as you are able to think about Fred, this will continue to be the case. You go home at night. (e4) is now replaced with:

(e5) there is some x such that, several hours ago, at a party, x was talking about the national debt, and x looked puffy and unhealthy, and "Fred" names x.

This process can continue *ad infinitum.* One's ability to think about, and refer to, something can only be *started up* by one's acquiring knowledge of the right kind of existence claim. Typically, that specific claim is forgotten. But it is replaced by another "interlocking" existence claim – another interlocking uniquely individuating description. And that second one will, in its turn, be replaced by a third; and that third by a fourth, and so on.

The point is that the ability to think about, and refer to, an external object always involves possession of a uniquely individuating description that

applies to it.

The objector says that there can be “pure”, i.e. non-descriptive, conceptions of objects. But conception is surely not so innocent; and this becomes clear when we consider the phenomenon of *recognition.* To recognize something is not, at least not merely, to register the truth of a bare identity proposition – a proposition of the form *O=O.* To register the truth of *Smith is Smith* is not, by itself, to *recognize* anything, and to recognize something is *not* (merely) to register the truth of *O=O.* Recognition is obviously replete with knowledge of descriptive content. When you recognize the man in the distance as the man who insulted you at the restaurant, a great deal of *descriptive*, and therefore existential, knowledge is involved. You know that *there was some x such that x insulted me at the restaurant.*…You also see that *there is some y in the distance such that*…Your *recognizing* the man in the distance *as* the man from the restaurant involves your “integrating” or putting together this information: you see that *both* descriptions – *man in the distance* and *man from the restaurant* – connect somehow; you see that there is some one thing such that *both* descriptions apply to it. In other words, you recognize the truth of some existential claim like: there is some x such that *x is a man who insulted me at the restaurant and is also a man in the distance whom I am now seeing.* Once it is granted that recognition has a descriptive, and therefore an existential, component; and once it is granted that recognition involves putting together this existential-descriptive information; it becomes highly plausible to see recognition as consisting in one’s registering the truth of the kind of existence-claim described a moment ago. The alternative is to see recognition as stripped of any descriptive content: but this is not psychologically (or, I think, epistemologically) plausible.

Let us relate these points about recognition to our analysis of conception –

the two things are closely related. When you *recognize* something, what is happening is that two *conceptions* converge. Your *recognizing* the man in the distance as the man at the restaurant involves your putting together different pieces of descriptive information; it involves your realizing that some one conception (the awareness you are having now, of the man in the distance) and some other conception (the one you have of some lout whom you dealt with at a restaurant) both pick out some one thing. Now if conceptions were "pure", i.e. were non-descriptive and therefore non-existential, then the just mentioned convergence of conceptions would be stripped of any descriptive content. Descriptive information wouldn't be involved; and recognition would be reduced to the registering of some bare claim of the form *O=O*. But that is not what recognition is. Recognition is obviously replete with descriptive, and therefore existential, content; and given the essential role that conceptions have in recognition, it becomes necessary to ascribe some of this descriptive knowledge to conceptions. So it becomes necessary to see conceptions as being paved with descriptive-existential knowledge. This fact, combined with the fact that we often forget the description which *starts up* a conception, gives plausibility to the analysis we gave a moment ago regarding the sustaining of a conception: a conception is sustained by a series of interlocking existence claims.

None of this belittles the role that causal connections play in conception and reference. I am able to think about, and refer to, Fred precisely because I *saw* him: and seeing something involves being causally connected to it. But *how* is the causal connection involved? What enables me to think about Fred is my knowledge of some uniquely individuating description: *there is some guy whom I am now seeing and attending to, who is wearing a bowler hat, and "Fred" names that guy.* The causal connection is involved in my seeing Fred; and my seeing Fred, in its turn, is involved in my having a uniquely

individuating description of him. So what is doing the work is my knowledge of a uniquely individuating description of Fred; the causal connection, though essential, has a role that is subordinate and internal to my knowledge of that description.

The most important reason *not* to think of conception as "pure", i.e. as descriptively vacant, is that such a conception makes it impossible to deal with Hesperus/Phosphorous cases. And the most important strike in favor of our analysis is that it enables us to deal precisely with such cases – in a way that is Kripke-friendly (and Putnam-friendly) but also psychologically viable. Let us now turn to this.

Here we need to be very clear about a certain distinction: the distinction between *literal meaning* and *cognitive content.* "Fred" is just a label. The following statement is true:

(F1) There is somebody x who I met on such and such occasion, and who looked pale and talked about communism; and the semantic rule for "Fred" is: "Fred" names x.

The following statement is false:

(F2) The semantic rule for "Fred" is this: There is somebody x who I met on such and such occasion, and who looked pale and talked about communism; and "Fred" names x.

The following is even *more* false

(F3) The semantic rule for “Fred” is this: for any predicate *…x…*, *…Fred…* means: There is somebody O who I met on such and such occasion, and who looked pale and talked about communism and…O…

The literal meaning of

(*) “Fred looked puffy and pale”

is simply:

(*LM) *Fred looked puffy and pale.*

But remember that one knows what “Fred” refers to only in so far as one knows some claim of the form: *somebody x uniquely has phi and “Fred” refers to x.* For one to know what “Fred” refers to *is*, or at least necessarily involves, one’s knowing some such claim. When you hear a sentence like (*), you have to interpret it. There is a difference between hearing the sounds and knowing their meaning. To interpret it, you have to exploit your background knowledge. You know what “Fred” refers to only in so far as you know some claim of the form: *somebody x uniquely has phi and “Fred” refers to x.* To interpret (*), you have to exploit that knowledge. So even though the literal meaning of (*) is simply (*LM), nonetheless you *access* that literal meaning only by way of some claim of the form: *somebody x uniquely has phi and “Fred” refers to x.* So what (*) *conveys* to you – not what it literally means, but what it conveys to you (what it means *to you*, we might loosely put it) – is some claim of the form:

(*CC) somebody x uniquely has phi and “Fred” refers to x, and x looked

puffy and pale.

Here we have to be very careful if we are to stay on the right side of Kripke's discoveries.

The sentence

(fa) "I am looking at Fred and he looks puffy and pale,"

is *not* analytic; it doesn't matter how you added the word "Fred" to your lexicon. Everybody who speaks English knows that (fa) is not definitionally true (or otherwise analytic). Everybody knows that (fa), though true, *could* have been false. Now suppose somebody points to a puffy, pale person and says "that guy is named 'Fred'". In that case, as we've seen, "Fred" is being added to your lexicon by way of your learning some existence claim like:

(pp) there is some x such that I am looking at x and x looks puffy and pale, and "Fred" refers to (names, means) x.

Of course, your friend did not explicitly *say* pp; but pp was the import of his ostensive definition; speaking very loosely, we might say that it was the implicature borne by his words. Notice that, in that implicated message, the semantic operator ("names", "refers to", "means") is given the narrowest possible scope. Because of this, you know that what (fa) literally means is some proposition of the form

(x) O looks pale and puffy,

for some individual O. And, for the same reason, you know that (fa) does *not* encode some existence claim.

Thus you will have no temptation to regard as analytic sentences like (fa); you will know that (fa) encodes something of the form *O looks pale and puffy* and, thus, that it does *not* encode something like *anything pale and puffy is pale and puffy* or any other analytic proposition.

Let us sum up. Whenever you add a referring term N to your lexicon, you do so by way of learning the truth of some claim of the form: something x uniquely has phi and N names (refers to, means) x. The semantic operator is given the *narrowest possible scope*. That is the reason why "N has phi" is *not* analytic. Further, at some level, you *know* that it must be given the narrowest possible scope. That is the reason you will rightly judge "N has phi" *not* to be analytic. Given *any* predicate *…x…* -- with some trivial exceptions – the preceding discussion makes it clear why you rightly judge *…Fred…* not to be analytic.

The exceptions I have in mind are those in which "Fred" occurs only vacuously or trivially, for example: "thing x such that x is either flat or not flat", "extended object", "thing x such that 1+1=2". Let us describe such predicates as "vacuous". (Our usage of this term is related to, but distinct from, the usage it has in logic.) A "vacuous" predicate is one that applies to any spatiotemporal entity. So for any non-vacuous predicate *…x…*, our discussion makes it clear why *…Fred…* is not analytic, and also why anyone who knows the semantics for "Fred" rightly judges it not to be analytic.

VI.

Kripke pointed out that “Aristotle was a great philosopher”, “Napoleon was a great general”, and the like are not analytic. Given what we just said, it is clear why our analysis is consistent with this fact, and why it is on the right side of Kripke’s well substantiated anti-descriptivism.

At the same time, our analysis explains why “Hesperus is Phosphorous” and “Fred is identical with Tommy” can *communicate* non-trivial propositions. We already explained why (fa) communicates (pp).

Now suppose that a few days later, you are at another party; this time it is a costume party. Somebody is dressed up in a suit of armor. Your companion happens to be acquainted with that person; but he knows him only by his street name, which is “Squiggy”. In this case, the word “Squiggy” is being added to your lexicon. For reasons that we’ve covered, that addition is effected by way of your learning some true claim of the form:

(^^) something x uniquely has phi and “Squiggy” refers to (names) x.

The actual claim in question is something like:

(^^^) Right now, there is somebody x that I am attending to, and x is wearing a suit of armor, and “Squiggy” names x.

The *literal meaning* of

(s) “Squiggy is drunk”

is simply some claim of the form

(s2) *O is drunk.*

But for reasons that we've discussed, what (s) *communicates* to you is not a bare ascription of drunkenness to some individual. Under the circumstances, what (s) *communicates* to you – what it means *to you* – is some existence claim, something of the form:

(s3) Right now, there is somebody x that I am attending to, and x is wearing a suit of armor, and "Squiggy" names x, and x is drunk.

Now we've already established that the literal meaning of

(f1) "Fred is drunk"

is simply:

(f2) Fred is drunk.

But we've also established that what (f1) *communicates* to you is some existence claim. The existence claim which it communicates is a function of the peculiarities of the circumstances under which you added "Fred" to your lexicon. If *those* circumstances are sufficiently different from the circumstances under which you added "Squiggy" to your lexicon, then the propositions that (f1) and (s1) will communicate to you (not the ones they literally mean) will be radically different; the one needn't entail the other. (f1) might communicate the claim:

(f3) there is somebody x whom I saw at a party ten days ago and x looked puffy and pale and "Fred" names x, and x is drunk.

Of course, (f3) doesn't entail, and isn't entailed by, (s3).

It is pretty clear how to apply to this to the Hesperus-Phosphorous case. You learn what "Hesperus" means through some existence claim:

(h1) there is some lovely celestial object x such that x is the only body to have yet appeared in the evening sky and "Hesperus" names x.

"Phosphorous" is added to your lexicon through a different existence claim:

(p1) there is some lovely celestial object x such that x is the last body to remain in the morning sky and "Phosphorous" names x.

A consequence is that

(h2) "Hesperus is lovely"

communicates to you (not what it literally means) is

(h3) there is some lovely celestial object x such that x is the only body to have yet appeared in the evening sky and "Hesperus" names x, and x is lovely,

whereas the proposition that

(p2) "Phosphorous is lovely"

communicates to you is

(p3) there is some lovely celestial object x such that x is the last body to remain in the morning sky and “Phosphorous” names x, and x is lovely.

Thus the propositions that (h2) and (p2) *communicate* to you are entirely different; neither entails the other.

But, in (h1) and (p1), the semantic operators are given the narrowest possible scope. So “Hesperus” is just a label – not a description quantifier – and the same is true of “Hesperus”. Thus for any non-vacuous predicate *…x…* -- including “celestial body that appears before any other in the morning sky”, and “celestial body that disappears after any other from the evening sky” -- what *…Hesperus…* literally means is non-analytic; the same is true of what *…Phosphorous…* literally means. Further, as a competent speaker of English, you *know* (if only tacitly) that the semantic operators in question are to be given narrow scope; so you will rightly judge that *…Hesperus…* and *…Phosphorous…* are non-analytic, for all non-vacuous choices of *…x…*.

So everything works out exactly as we want it to. We have a semantics that is consistent with Kripke’s well confirmed point that names are not descriptions. But notwithstanding that fact, we have a pragmatics (or an epistemology, perhaps) that explains why *…Hesperus…* may *communicate* something different from *…Phosphorous…*. Given that “Hesperus” and “Phosphorous” are not descriptions, it is hard to believe that they are anything other than mere names, mere labels. Given that they name the same thing, it follows that (h2) and (p2) have the same proposition for

their literal meanings. This is all fine. But we need to reconcile *that* fact with the fact that (h2) and (p2) communicate different propositions. This reconciliation is not so hard to make; we simply need to identify the facts about how we *learn* semantics. The difference in cognitive value between (h2) and (p2) is not to be explained semantically; it is not to be explained by tinkering with the semantic rules for "Phosphorus" and "Hesperus"; it is to be explained by carefully identifying the facts surrounding our coming to learn those rules.

VII.

If you accept the thesis that "Hesperus" and "Phosphorous" are mere labels, then (supposedly) you have a devil of a time accounting for the obvious cognitive difference between (h2) and (p2). At the same time, given what Kripke said, nobody wants to deny that these names are mere labels. This is quite a puzzle: let us refer to it as the Direct- Reference Theorist's Dilemma (DRTD).

Some have dealt with DRDT by advocating a "two-dimensionalism" about names. Different *tokens* of "you" refer to different things. The *type* "you" does not refer to anybody. But that type still has a meaning: it is a function that assigns referents to tokens of that word, on the basis of facts about the context of utterance. This function is the *character* of the type "you". A given *token* of "you" has content: this is the thing to which that token refers. (So if I say "you are tired" to Smith, the content of that token of "you" is Smith.) Thus "you", and indexicals generally, have a "two-dimensional" semantic structure.

The character of "you" is associated with descriptive content (*addressee of the speaker's words*). The *content* of a token of you has no descriptive

content; it is simply an individual. But because indexicals have a two-dimensional structure, they can be both directly referential *and* have descriptive content.[267]

Some have dealt with DRDT by saying that proper names, like indexicals, have a two-dimensional semantic structure. So tokens of "Hesperus" and "Phosphorous" are indeed mere labels. But the *types* corresponding to those tokens have characters, or something similar to characters; and those characters are associated with different descriptive contents. And the difference in *cognitive* value between (p2) and (h2) is rooted in the fact that "Hesperus" and "Phosphorous" have different characters; in any case, they differ with respect to some character-like semantic dimension.[268]

§ I don't think that this two-dimensionalism is necessary to account for that cognitive difference. (I also don't think it is an available option.) We've seen that reference is always *fixed* by descriptive information. The description is given *wide* scope (with respect to the semantic operator). Because of that, the name in question is a mere label. But given only that *semantically* the descriptive information takes wide-scope, it doesn't follow that it is *cognitively* inert. You will still *associate* that descriptive information with the name. You cannot access word meaning in a vacuum. The meaning of "Hesperus" is simply Hesperus (Venus/Phosphorous). But there is a difference between hearing "Hesperus" and assigning it the right meaning. Even though the meaning of "Hesperus" is simply Hesperus, your *knowing* that fact consists in your knowing some proposition of the form: *something x uniquely has phi and "Hesperus" names x.* To assign the right referent to a token of "Hesperus" you must *exploit* this knowledge. Thus, inevitably, sentences of the form "…Hesperus…" will *communicate* something

descriptive, something existential, even though what they semantically encode has the form: …*O*…

Differences in cognitive value don't always have a strictly *semantic* basis. Sometimes they do – but not always. The sentence "snow is white" has a different cognitive value from "grass is green"; and that fact has a semantic basis. But that is not always how it works. Suppose you look through a microscope at Alpha, and you are told by your lab partner*: there is some x such that you are looking at x, and let "Smith" refer to x and let "Smith" have no other semantic content.* Later, without your knowledge, you are again looking at Alpha, this time with the naked eye, and you are told: *there is some x such that you are looking at x, and let "Jones" refer to x and let "Jones" have no semantic content other than x.* By our stipulation, "Smith" and "Jones" have exactly the same semantic content. So "Smith is lovely" and "Jones is lovely" have the same proposition for their literal meaning. But, also obviously, they have different cognitive values. This is because reference cannot be fixed, or accessed, in a vacuum. Even though "Smith" is a mere label, the act of putting that label on a thing, and the act of figuring out which thing that label is on, is necessarily paved with descriptive information. The semantics of an expression is one thing; *figuring* out its semantics is another thing. Instead of taking refuge in dubious two-dimensionalism, we need only distinguish semantics proper from the epistemology of semantics.

VIII.

It has often been said that to think about an object, one has to be causally

connected to it. This has been held both by semanticists and also by philosophers of mind. Kaplan (1968) held it early in his career (though he later retracted it); so did Donnellan. Much of Fodor's work is devoted to arguing that to be aware of an object – to have a concept of it – just *is* to stand in a certain kind of causal relation to it.

This view *seems* to conflict with the fact that we can refer to things to which we have no causal connection. The famous example is "Newman 1", which is defined thus: if there is an x such that is borne before anyone else in the 3rd millennium A.D., let "Newman 1" refer to x. So if Brown has that property, then

(N1) "Newman 1 will be a genius"

semantically encodes the proposition:

(BG) *Brown will be a genius.*

It seems pretty clear that we can *make* our expressions refer to anything we want; it doesn't matter whether we have the right causal liaisons or not. But, for whatever reason, many have also held that one must be *causally connected* to a thing to think about it. They have dealt with this situation by saying that we don't grasp the proposition semantically encoded in (N1); we are not causally connected to Brown (at least not in the right way); therefore we cannot think about him; therefore we cannot assign the right referent to "Newman 1". For in order to do that, we'd have to grasp Brown; but we cannot, owing to our not having the right causal relation to him.

Given what we've said in this paper, this reasonable-seeming view should

be reconsidered. It is undeniably true that, in many cases, our ability to think about a thing involves our having a certain causal relation to that thing. But we've seen that, in every such case, the causal connection between subject and object is simply an ingredient in one's knowledge of the right kind of uniquely individuating description. In *all* cases, one's conception of a thing consists in one's knowing some existence claim of which that thing is the sole verifier. If you acquire a concept of Fred by *seeing* him (or otherwise sense-perceiving him), that consists in your acquiring knowledge of an existence claim of which Fred is the unique verifier. If you acquire a concept of Fred by being *told* about him, or by hearing about him, that too consists in your acquiring knowledge of an existence claim of which Fred is the unique verifier. In all cases, then, knowledge of a thing consists in knowledge of a uniquely individuating description that applies to that thing. In some cases, a causal connection between subject and object is an ingredient in that knowledge. But it is never the causal connection *per se* that supports conception; it is always knowledge of a uniquely individuating description.

If I am able to think about Socrates, it is – we have found, in direct contradiction to the contemporary orthodoxies – entirely because I know some uniquely individuating description that applies to him. Conception is knowledge of a uniquely individuating description. The presence of a causal relation between subject and object, when present, is simply a means to that end. Obviously we can know of a uniquely individuating description that applies to Brown (Newman1). So it seems wrong, or at least arbitrary, to say that we *can* think about Socrates but that we cannot think about Newman 1. For the very thing that enables us to think about Socrates is present in the case of Newman 1.[269]

Bibliography

Meaning and Truth. 1991. Jay L. Garfield and Murray Kitely (eds.)
St. Paul, Minnesota: Paragon House

The Philosophy of Language. (1990.) Second Edition A.P. Martinich (ed.).
Oxford: Oxford University Press.

Bach, Kent. 1984. *Thought and Reference.* Oxford: Clarendon.

Berkeley, George. 1934. *An Essay Towards a New Theory of Vision.* Oxford: Clarendon Press.

Blackburn, Simon. 1984. *Spreading the Word.* Oxford: Clarendon.
--1998. *Ruling Passions: A Theory of Practical Reasoning* New York: Oxford University Press.

Burge, Tyler. 1979. "Individualism and the Mental". In French, Uehling, and Wettstein (eds.) *Midwest Studies in Philosophy* IV, Minneapolis: University of Minnesota Press, pp. 73-121.
--1982. "Other Bodies". In Woodfield (ed) *Thought and Object.* Oxford: Oxford University Press.
-- 1994. "Individualism and Self-knowledge". In *Self-Knowledge.* Edited by *Quassim Cassam.* New York: Oxford University Press.

Cain, M.J. 2002. *Fodor: Language, Mind and Philosophy.* Cambridge: Polity Press.

Davidson, Donald. 2004. “Turing’s Test”. In *Problems of Rationality.* Oxford: Oxford University Press.

Donnellan, Keith. 1990/1966. “Reference and Definite Descriptions.” In A.P. Martinich (ed). *The Philosophy of Language.* Oxford: Oxford University Press.

Evans, Gareth. 1990/1974. “The Causal Theory of Names”. In A.P. Martinich (ed). *The Philosophy of Language.* Oxford: Oxford University Press.
-- 1980. “Pronouns”. *Linguistic Inquiry* 11: 337-362.
-- 1982. *The Varieties of Reference.* Edited by John MacDowell. Clarendon: Oxford University Press.
-- *Collected Papers.* 1985. Oxford: Clarendon Press.
-- “Reference and Contingency”. 1985. In Evans 1985.

Field, Hartry. 1977 “Logic, Meaning, and Conceptual Role” *Journal of Philosophy* Volume LXXIV, Number 7, July 1977.

Fodor, Jerry. 1975. *The Language of Thought.* New York: Thomas Y. Crowell
--1981a. “Introduction”. In *Representations.* Cambridge: The MIT Press.
--1981b. “Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology”. In *Representations.* Cambridge: The MIT Press.
--1981c. *The Present Status of the Innateness Controversy.* In *Representations.* Cambridge: The MIT Press.

--1987. *Psychosemantics.* Cambridge: The MIT Press.
-- 1990a "Fodor's Guide to Mental Representation". In *A Theory of Content and Other Essays.* Cambridge. The M.I.T. Press.
-- 1990b. "A Theory of Content I". In *A Theory of Content and Other Essays.* Cambridge. The M.I.T. Press.
--1990c. "A Theory of Content II". In *A Theory of Content and Other Essays.* Cambridge. The M.I.T. Press.
--1990d. "Substitution Arguments and the Individuation of Beliefs". In *A Theory of Content and Other Essays.* Cambridge. The M.I.T. Press.
--1994a. *The Elm and the Expert: Mentalese and Its Semantics.* Cambridge: M.I.T. Press.
--1994. "Fodor". In *A Companion to the Philosophy of Mind.*(Samuel Guttenplan ed.). Cambridge: Blackwell.
--1998. *Concepts.* Clarendon Press: Oxford.
-- 2001. *The Mind Doesn't Work that Way.* Cambridge: M.I.T. Press.

Frege, Gottlob.1990/1982a. "On Sense and Nominatum." In A.P. Martinich (ed) *The Philosophy of Language* Oxford: Oxford University Press.

Hale, Bob and Wright, Crispin (eds.). 1997. *A Companion to the Philosophy of Language.* Oxford: Basil Blackwell.

Horst, Stephen. 1996. *Symbols, Computation, and Intentionality.* Berkeley: University of California Press.

Kaplan, David "Quantifying In". 1990/1968. In A.P. Martinich (ed). *The Philosophy of Language.* Oxford: Oxford University Press.

--1990/1970. “Dthat”. In A.P. Martinich (ed). *The Philosophy of Language.* Oxford: Oxford University Press.
-- 1978. “On the Logic of Demonstratives”. *Journal of Philosophical Logic* 8 (1978): 81-98.
-- 1989. ‘Demonstratives’. In *Themes from Kaplan.* Edited by Joseph Almog, John Perry, and Howard Wettstein. Oxford: Oxford University Press.
-- 1989b. “Afterthoughts”. In *Themes from Kaplan.* Edited by Joseph Almog, John Perry, and Howard Wettstein. Oxford: Oxford University Press.

Kripke, Saul. 1999/1971. “Identity and Necessity”. In *Metaphysics: An Anthology*. Edited by Jaegwon Kim and Ernest Sosa. Oxford: Blackwell.
--1972. *Naming and Necessity.* Cambridge: Harvard University Press.
-- “A Puzzle about Belief”. 1979. In *Meaning and Use,* ed. A. Margalit. Dordrecht: Reidel.

Lepore, Ernie. 1994 “Conceptual Role Semantics” in *A Companion to the Philosophy of Mind* edited by Samuel Guttenplan Blackwell Cambridge, MA.

Lewis, Clarence. 1946. *An Analysis of Knowledge and Valuation.* Cambridge: Harvard University Press.

Loar, Brian (1988) “Social Content and Psychological Content” in Grimm and Merrill (eds.) *Contents of Thought* University of Arizona Press.

Loewer, Barry. 1999. ‘A Guide to Naturalizing Semantics” in *A Companion to the*

McGinn, Colin. 1988. *Mental Content.* Oxford: Oxford University Press.

Mill, John Stuart. 1991. “Of Names and Propositions”. In Jay Garfield and Murray Kiteley (eds.) *Meaning and Truth.* St. Paul, Minnesota: Paragon House.
-- 2002. *A System of Logic.* University Press of the Pacific.

Perry, John. 2000. *The Problem of the Essential Indexical and Other Essays.* Palo Alto: CSLI Publications.
Putnam, Hilary. 1975. “The Meaning of ‘Meaning’”. In *Collected Papers: Volume II.* Cambridge: Harvard University Press.

Rosenthal, David M. (ed.). 1991 *The Nature of Mind.* New York: Oxford University Press.

Russell, Bertrand. 1903. *Principles of Mathematics.* Cambridge: Cambridge University Press.
--1990/1905. “On Denoting”. In A.P. Martinich (ed) *The Philosophy of Language* Oxford: Oxford University Press.
-- 1957/1910. “Knowledge by Acquaintance and Knowledge by Description". In *Mysticism and Logic.* Garden City, New Jersay: Doubleday.
-- 1920. *Introduction to Mathematical Philosophy.* London: George Allen Unwin.
--1921. *The Analysis of Mind.* London: George Allen Unwin.
-- 1948. *Human Knowledge: Its Scope and Limits.* London: George Allen

Unwin.

Salmon, Nathan. 1986. *Frege's Puzzle.* Cambridge: M.I.T. Press.

Soames, Scott. 2001. *Beyond Rigidity.* Princeton: Princeton University Press.

Stich, Stephen P and Warfield, Ted A (eds.) 1994 *Mental Representation: A Reader* Blackwell Oxford: United Kingdom

Strawson, Peter. 1990/1950. 'On Referring'. In A.P. Martinich (ed.) *The Philosophy of Language.* Oxford: Oxford University Press.
--1959. *Individuals.* New York: Routledge.

Stroud, Barry. 2001. *The Quest for Reality.* Oxford: Oxford University Press.

Intentionalism and Foundationalism

Abstract: Intentionalism is the doctrine that mental states cannot differ in phenomenal content unless they differ in respect of what they represent: phenomenology "supervenes" on representational content. Intentionalism is thus committed to the view – in fact, it *is* the view – that a given kind of phenomenal content *intrinsically* has a certain representational content: nothing must be *added* to that kind of phenomenology to associate it with that kind of representational content; no contextualization, no back-ground knowledge, no ulterior factors are needed to secure that association. (Compare: given that the aesthetic properties of the statue supervene on its physical properties, it follows that nothing must be *added* to those physical properties to secure the connection between those physical properties and those aesthetic properties.) Wilfred Sellars argued compellingly that what makes a given bit of phenomenal content have a given kind of representational content is *precisely* the addition of such background knowledge – of such contextualizing information: to think otherwise, said Sellars, is to be guilty of "the myth of the given" – a spurious epistemic foundationalism. In this paper, I merely bring out the connection between intentionalism and the myth of the given. I argue that, unless Sellars is completely wrong, intentionalism must be completely wrong.

The purpose of this paper is to show that a doctrine known as *intentionalism* is guilty of what Wilfred Sellars called 'the myth of the given', and is therefore untenable.

What is intentionalism? Some mental entities have both phenomenal

content (phenomenology) and representational content. My perception represents the world as being a certain way; so it has representational content. And there is 'something it is like' to have that perception; so it has phenomenal content (it has 'phenomenology').

Intentionalism says that two experiences cannot have different phenomenal contents *unless* they have different representational contents. So phenomenal content 'supervenes' on representational content.

I submit that intentionalism is guilty of the 'myth of the given'. Let me say what this means by way of a Sellarsian story.

You are in a tie shop in which the lighting conditions are a bit unusual. [270] You are looking at a tie in a display case. The tie induces various visual sensations in you. On the basis of these sensations, you can't quite tell whether the tie is orange or red or maroon. You take the tie outside to examine it in better lighting. You see that it is maroon. Now you know what a maroon tie looks like under the lighting conditions that obtain in the tie shop. So the next time you see what is in fact a maroon tie in the tie shop, you *don't* have to take it outside to determine that it is maroon.

But if a tie is some color other than maroon– say, burgundy or maroon -- you cannot know that on the basis of the visual sensations you have in the tie shop – for the same reason that, earlier, you couldn't tell that a tie was *maroon* on the basis of the sensations you had in the tie shop. Let me illustrate my meaning with a new story. You are again in the tie shop, and are looking at what is in fact a burgundy tie. You can't tell what color it is. So you take it outside. There you *see* that it is burgundy. You take it back inside and attend to its appearance. You then know that *that* is what a burgundy neck-tie looks like in the conditions that obtain in the tie shop.

You repeat this process over and over again, until you can tell *for any color C* what an object with C looks like in the lighting conditions that obtain

in the tie shop. The correspondence that obtains between objects' colors *in the tie* shop and one's visual sensations is different from the correspondence that obtains between objects' colors and one's sensations outside the tie-shop. By repeating the processing described a moment ago, you learn the former correspondence-rule. And you no longer have to go outside to know what color a tie is. After a while, if you manage to internalize this rule of correspondence, you will, while inside the tie-shop, be able to just *see* what color a tie is. No doubt somebody who has been working in the tie-shop a long time can *see* what colors the ties have; he doesn't have to go through a laborious process of inference (or, at least not a conscious one).

What does this tell us? When you are in the tie shop, your visual sensations do not *by themselves* tell you what color a given object is. What tells you this is your visual sensation *plus* knowledge of various environmental conditions that obtain. By itself, your visual sensation is color-indeterminate. What is color-determinate is the visual sensation *plus* a great deal of background knowledge.

So, when you are in the tie-shop, your visual sensations (at least so far as they relate to color-facts) are not 'foundational'; in order for them to tell you what colors objects are, they must rely on ulterior knowledge that you have: their ability to tell you anything *presupposes* knowledge, on your part, of various environmental conditions.

Let us sum up. In some contexts, your sensations do not *by themselves* tell you what colors objects are; only those sensations coupled with background knowledge about the conditions of perception tell you that.

It is easy to generalize this argument – to make it apply to *any* perceivable property P. (By a 'perceivable' property, I mean one that we *sense*-perceive objects to have. We don't sense-perceive that an object consists of atoms with three electrons. So that property is not perceivable. But we do sense-perceive

that an object is more than ten feet away from us. So that is a perceivable property.[271])

Ordinarily we think that we can just *see* how distant an object is from us. But when I just *see* that a certain tree is less than a hundred feet from me, a lot of background information is at work. For example, you take it for granted that the tree is not 10,000 feet tall. If you thought that the tree *was* 10,000 feet tall, then a phenomenally identical perception would lead you to think that you were seeing a tree was miles away, as opposed to 80 ft away. Let us amplify and substantiate these remarks through a Sellars-type thought experiment.

When you look through a microscope, small things appear big. When you look through a telescope, big things appear small. Imagine that, every morning, you woke up wearing a pair of goggles that either magnified or diminished the strength of your sight. On Monday, you wake up wearing a pair of microscope-goggles – so it would be as though you were looking through a microscope. Another day, you wake up wearing a pair of telescope goggles – so it would be as though you were looking through a telescope. And everyday it would be different; the magnifying power of the goggles would be different every day. (So one day it would be as though you were looking through a weak magnifying glass. Another day, it would as though you were looking through a moderately powerful telescope.)

Under these circumstances, your visual perceptions at the start of any given day would be distance-indeterminate and, therefore, size-indeterminate. You wouldn't know whether you were seeing an object that was 100 feet tall and a light-year away or an object that was ten feet tall and was ten feet away. You would have to *learn* what correspondence obtained between the character of your visual perceptions, on the one hand, and the distances and sizes of objects.

One way you could do this is as follows. You look at a given object; you don't know whether it is a ten foot object that is 100 ft off or a ten-thousand foot object that is ten miles off. You take off the goggles and you simply *see* how far off, and how tall, the object was; you *see* that it is a building that is approximately 100 ft tall that is 100 ft away. You then put the goggles back on and, now *knowing* the distance and size of the object, you look at it again: you know now that an object that, when you are wearing the goggles, looks like *that* is really an object of such and such dimensions and thus such distance away from you. You repeat this process, with objects of other sizes and distances. And eventually you learn the correspondence that obtains between, on the one hand, the character of the visual perceptions that you have while wearing the goggles and, on the one hand, the distances and sizes of the objects you are seeing. Eventually, you will, while wearing the goggles be able to simply *see* that the object your are looking at is (say) ten feet tall and a 100 ft away – just as somebody who works in the tie-shop can simply *see* that a given tie is burgundy, even though the lighting conditions are non-standard. But, of course, when you simply *see* (while wearing the goggles) that a given object is ten ft tall and a 100 ft away, it is not your visual sensations *by themselves* that apprise you of the object's size and distance; what really does this is your visual sensations *plus* a great deal of back-ground knowledge about the conditions of perception.

It is important to note that, before you learn the just mentioned correspondence, your visual sensations are distance- and size-indeterminate. A given sensation could tell you *either* that an object is 10,000 ft tall and 100,000 ft away or that it is 20 ft tall and 10 ft. away. Until you acquire the relevant background knowledge, your visual sensations are simply *indeterminate* in these respects; your visual sensations don't tell you *either* that the object is 10,000 ft tall and a 100,000 ft away or 20 ft tall and 10 ft

away. In response to this, one might say:

Under these circumstances, your visual sensations *would* give you determinate messages, but these messages would be wrong. You wake up in the morning wearing goggles that, let us suppose, quadruple the strength of your vision. In that case, your visual sensations give you systematically false messages: you *see* objects as being four times larger than they are and, thus, four times closer than they actually. So it isn't that your visual sensations are size- and distance-indeterminate; it is that they *are* size- and distance-determinate but that they are determinate *in the wrong way.*

At first this seems like a reasonable view. But really it is very arbitrary. As Tyler Burge[272] said, a perception to the effect that P is necessarily *prima facie* evidence that P. If you *see* a red chair in front of you, that is necessarily prima facie evidence that there is a red chair in front of you. In the story we are telling, you wake up every morning with goggles of a different level of magnification attached to your eyes, and you never wake knowing what the level of magnification is. Under those circumstances, your initial sensations are not even *prima facie* evidence of facts relating to the sizes or distances of objects. Given only your initial sensations, you would be equally justified in believing that you were seeing 100 ft object 100 ft away or a ten thousand foot object 10,000 feet away. So your initial sensations wouldn't be *prima facie* evidence of either state of affairs – or, by analogous reasoning, of any other state of affairs relating to the distance or size of the object in question. Thus your initial perceptions wouldn't *tell* you that the object was this as opposed to that size or distance; you wouldn't, veridically or non-veridically, *see* the object as being of this or that size or distance.

The objector's point can be confuted in another way. Again, suppose you wake up wearing goggles that strengthen your vision four-fold, but when you wake up you don't *know* what magnification these goggles have. After you 'acclimatize' to your goggles, you will simply *see* objects as having the dimensions that they in fact. (We are leaving aside cases of objects that are extremely distance or extremely small. One doesn't really *see* a star to a million light years away; one doesn't *see* it, veridically or non-veridically, to be *any* determinate distance away from one. And of course one doesn't *see* a proton, veridically or non-veridically, to have this or that size; for one doesn't see protons at all.) After this acclimatization, when you look at an object that is ten ft tall and twenty ft away, you won't *first* see it as forty ft tall and eighty ft away; you will see it for the size and distance that it actually has – just as the veteran of the tie-shop, even while in the tie-shop, will *see* ties for the colors they actually have: he won't *see* a burgundy tie as being orange and *then* consciously figure out, on the basis of his background knowledge, that what he is seeing is *in fact* burgundy. He will *see* the tie for the color it has.

Let us refer to the person who has acclimatized to his goggles (with magnification × 4) as Post-Bob, and let us refer to that same person, after waking up but prior to that acclimatization, as pre-Bob. We've agreed that Post-Bob *sees* objects as having the sizes and distances that they actually have. (That is, Post-Bob has size-right and distance-right perceptions *in so far as any one does*. Whether we're dealing with Post-Bob or with a 'normal' person of 20/20 vision, there will inevitably be perceptual vagueness and error, especially where extreme sizes or distances are concerned). Now the objector says that pre-Bob simply *sees* objects as being 4 times their actual size: so his perceptions are not size-indeterminate, but systematically size-*wrong.* But that puts the objector in an awkward position. Pre-Bob's visual

sensations don't change as he metamorphoses into Post-Bob. But it is very hard to see how a determinately *false* perception could become a determinately *correct* perception simply as a result of a change in background knowledge. It seems that if, when you woke up, you had perceptions that were systematically *false*, as opposed to indeterminate, a change in background knowledge wouldn't result in *veridical* perception: it would simply enable you to intellectually [273] *correct* for your *non-veridical* perceptions. But this isn't what happens: Post-Bob *sees* things the right way; he doesn't *see* things the wrong way and then make a mental note to himself along the lines of 'I see that tree as being 80 ft tall, so it must really be 20 ft tall'. No – he *sees* the tree as being the height that it *actually* is.

So it is hard to see how systematically *false* perceptions could be turned into veridical ones through a change in background knowledge. But it is easy to see how *indeterminate* perceptions can be turned into determinately *veridical* ones once background knowledge is filled in. If one's perceptions are indeterminate (in respect of, say, size or distance), then there is a kind of openness to correction built into them: there is a parameter in it that has yet to be set and that consequently *can* be set. But if a perception is determinately *false*, it is frozen; background knowledge won't turn it into a veridical one; the perception will continue to give one its wrong message (though, armed with the right background knowledge, one will be able to intellectually superimpose the right message). So it is very hard to see how a determinately *wrong* (or, to use Quine' word, a *falsidical*) perceptions can become veridical through a change in background knowledge. But it is very easy to see how a change in background knowledge can turn indeterminate perceptions (for example, size- or distance-indeterminate perceptions) into determinate ones. We are taking the latter position. The objector is committed to the former. Therefore the objector is wrong. Pre-Bob's perceptions are size- and

distance-*in*determinate, not size- and distance-*false.*

Let us sum up thus far. Under non-standard conditions of perception, one's sensations do not by themselves apprise one of the sizes or, therefore, the distances of objects. What does this is those sensations operating *in conjunction* with a great deal of background knowledge. Absent this back-ground knowledge, one's visual sensations are simply size- and distance-indeterminate.

Exactly similar reasoning shows that, under non-standard conditions, our perceptions are mass-indeterminate, temperature-indeterminate, velocity-indeterminate, and so on. Given *any* perceivable property P, under non-standard conditions, our sensations do not by themselves tell us to what degree, if any, x has P; only our sensations operating in conjunction with background knowledge (about the conditions of perception) do that. Under non-standard conditions, our sensations, taken by themselves, are temperature-indeterminate, mass-indeterminate, and so on. They become temperature- and mass-determinate *only* in so far as they are put in the context of a great deal of back-ground knowledge. In general, under non-standard conditions, perceptions are indeterminate. For any kind of perceptible property P – P could be size, distance, color…-- it would be possible to create non-standard perceptual conditions such that, under those conditions, one's perceptions were P-indeterminate. And, under such conditions, what makes one's perceptions become P-*de*terminate is the acquisition of *background* knowledge – knowledge of facts about the conditions of perception.

Another way of putting this is to say: for any kind of perceptible property P, it would be possible to create circumstances in which one's perceptions are P-determinate *only with the help of background knowledge.* Under such circumstances, when one saw an object O as having P (or as having P to

degree n), a great deal of background knowledge would be implicated in that perception. Consequently one's perception of O as having P would not be *foundational*; it would be built on top of, so to speak, a great deal of background knowledge about the conditions of perceptions. Under those circumstances, it would not be *given* to one that O was P; not even a *false* message to the effect that O was P would be given to one. One would be 'given' (if that is the right word) only a non-epistemic sensation; one's background knowledge would be needed to assign that sensation a P-value.

So far we've spoken about *non*-standard perceptual conditions. But it is easy to show that what we've said applies equally to standard perceptual conditions. For standard conditions are just non-standard conditions that one has got used to; and the distinction between standard and non-standard conditions is quite arbitrary. In fact, *no* conditions are quite standard; if we make conditions *perfectly* standard – the way they might be in a surgical theatre or in the kind of room where they make computer chips – the result is a situation where the perceptual conditions are so good that they are not standard any more. And one is *always* doing in a small way what, in the anecdote above, Bob was doing in a big way: one is always making allowances for background knowledge.

Let us once again think about Bob. One morning Bob wakes up fitted with goggles of magnification ×2. But this time the goggles stay on for years and years. This time, since Bob wouldn't be able to take off his goggles, it would not be an option for him to compare his goggle-produced perceptions with his non-goggle perceptions: so he couldn't' he would no longer have that way of figuring out *what* his visual modality was telling him. But he would still acclimatize. From his pre-goggle days, he obviously knows the sizes of objects – of stop signs and baseball bats, and so forth – and he could use these as benchmarks. Initially, he would *consciously* correct for his

perceptions by using these benchmarks; he see an object, not know how big it was, compare it to a baseball bat, see that it was the same size, and *then* know how big it was. But eventually, no doubt, Bob would simply *see* objects for the sizes they were. (At least in so far as anybody does: nobody's vision is without significant indeterminacy in many respects; one doesn't *see* very distant or very large objects as being specifically this or that size or, derivatively, *as* moving with this or that speed.)

Of course, in some ways, Bob's vision will be much better than that of an ordinary person's, and in other ways, it will be worse. (He will have superb far-sight, but terrible near-sight.) But this doesn't matter to our discussion. What matters is this. Bob's ×2 vision will *become* standard for Bob – just as your vision, with its ×1 magnification, is standard for you. It is clear, given what we've so said so far, that Bob's perceptions acquire determinate content (at least in respect of distance and size and related properties, like speed) only with the co-operation of background knowledge. Not only does Bob's ability to acquire *knowledge* (of distance, size, and related properties) presuppose a great deal of background knowledge. So during this ×2 phase of Bob's vision, sizes and distances (and related properties) would not be *given* to Bob; his perceptions would be size- and distance determinate only by courtesy of a great deal of background knowledge. So distances and sizes would be given to Bob in the sense that he could, within limits, see what sizes and distances objects had; they would be *made known* to him. But sizes and distances would not be given in the sense that his distance- and size-perceptions would be *foundational*; he would have size- and distance-perceptions *only* because there was a lot background knowledge at work. Of course, in so far as a hefty body of background knowledge must lend its help to make perceptions represent things as having determinate sizes and distances, to that extent sizes and distances are not *given* to one; one has to *earn* them, so to speak. So in a

special, but by no means arbitrary, sense of ‘given’, distances of and sizes are *given* to Bob during this ×2 phase of his visual career.

But now suppose that, after all these years, Bob’s revision reverts back to how it used to be. So it is no longer ×2, it is ×1. His vision now would *not* be standard for him, even though it used to be. He would have to do now what he had to do before when he first shifted to ×2 vision. He would acquire *new* background information, and this new background information would be implicated in making his *new* perceptions be size- and distance-determinate. So, just as distances and sizes were not *given* to Bob in his ×2 phase, so they are not given to Bob in his *new* ×1 phase. Exactly similar considerations apply.

But what about Bob’s *old* ×1 phase? Maybe, it might be thought, sizes and distances were just *given* to him then. But it would be highly arbitrary to say that. There are two reasons for this. First, it is very hard to believe that Bob’s *new* ×1 perceptions function in a completely different way from his *old* ×1 perceptions. We know that his new ×1 perceptions are size- and distance-determinate only by courtesy of a great deal of background knowledge concerning the conditions of perception. But if, during Bob’s *old* ×1 phase, distances and sizes were just *given* in perception (in our special sense of ‘given’), then the epistemic liason between Bob and the world constituted by Bob’s *old* ×1 perceptions would very different from that constituted by his *new* ×1 perceptions. But there seems no reason to believe this.

There is a second reason to think it arbitrary to say that during Bob’s *old* ×1 phase distances and sizes were *given* to him (in our special sense of ‘given’), whereas during his ×2 phase they are not given. We’ve been discussing extreme changes in the conditions of perception. But one’s visual career is ordinarily characterized by much smaller, but still very genuine, changes perceptual conditions. And what we said in connection with artificially

dramatic change in these conditions is just as true of the smaller versions of such changes that actually occur. One can scarcely walk ten feet, or even rotate one's head, without the lighting conditions changing. But usually (though not always) one doesn't have to make a conscious adjustment. One automatically makes the adjustments, and sees green objects *as* green and red objects *as* red.

The kind of sensation that mediates a perception of a particular shade of red is heavily dependent on highly variable conditions. One doesn't realize this until one tries one's hand at painting, and has to reckon with the setting of the sun or the uneven light cast by a lamp. To reproduce in painting two different objects that one *sees* to be of a single shade of red, one may have to use two very different shades of *paint.* When one looks at the paint*ing*, one sees the two objects as being of the same shade of red – just as the artist, when he was having the actual perception which he reproduced in the painting, saw the two objects as being the same shade. But when one looks at the *paint* – not the paint*ing*, not the things *depicted* by means of the paint – one sees two *different* shades of red. The differences in *shade* correspond to differences in *quale*, in *sensation.* The identity of color of the *objects* represented in the painting corresponds to the identity in representational content of the *perceptions* reproduced in the painting.

Thus one makes innumerable adjustments to ever changing perceptual conditions. So one is perpetually adjusting to non-standard conditions. We can only assume, by arguments analogous to those given earlier, that one's background knowledge is implicated in these perceptions. Colors (for example) are not *given* to one; what is given to one is a non-epistemic sensation – a sensation with no determinate color- (or distance- or size-…) content. What *does* constitute a color- (or size-…) determinate perception is that sensation *in conjunction with* a great deal of back-ground knowledge.

Or course, arguments exactly analogous to those that we've given in connection with size, distance, and color apply to mass, temperature, and any other kind of perceptible properties: perceptible properties are not *given* (at least not in our special sense of the word). Whenever one has a perception that is mass- or size- or temperature or color-determinate – that is, whenever one has a perception that even *purports* to tell one what color or temperature or mass… an object has – that is, in part, because one's quale *sensations* are interacting with a great deal of background knowledge. The sensations by themselves are not color- or size- or temperature-determinate. The sensations themselves – the 'raw feels' or quale -- are not representational. The sensations *plus* the background knowledge is what is representational.

Of course, a perception is, by definition, representational; a perception necessarily is either size- or color-… or mass-determinate; a perception necessarily gives one *some* kind of message and is thus not *wholly* indeterminate. So what we have seen is not, obviously, that *perceptions* are not representational (that would be absurd) or that facts about the world are not *given* through *perceptions* (that too would absurd: there is no other way that facts about the external world are given to us). What we have seen is that the very existence of any single perception *presupposes* the co-operation of background empirical knowledge. (This raises the question how we can ever have a *first* perception: a question handles very delicately. There isn't room to address this problem here. But I don't think it constitutes any significant obstacle to the anti-foundationalism sketched here.)

The 'myth of the given' is that a given sensation *by itself* constitutes some kind of awareness of an external reality, and that no background knowledge is presupposed. Of course, if this were the case, then sensations would be epistemic rock-bottom: they would constitute the foundations of all knowledge. In sensation, certain bits of information would be *given* to us – in

our special sense of ‘given’ – and all knowledge would be built on top of this given information. (And, in this context, to say that piece of information is ‘given’ to us in a sensation is to say that no *background* knowledge would be implicated in the acquisition of this information.)

Actually, what we’ve seen is not only that no *knowledge* is ‘given’ (in our special sense), but that nothing *representational* is given: background knowledge is implicated in the generation of anything that even *purports* to give information about the world. So not only is no *knowledge* given; *nothing* even representational is given. There are thus two myths of the given: one is that *knowledge* is given; the other is that there are perceptions *in a vacuum* of background knowledge – in other words, that in some cases one’s sensual experiences do not need the co-operation of background knowledge to constitute perceptions (or even misperceptions). I submit that intentionalism is guilty of the second ‘myth of the given’ and thus *a fortiori* is guilty of the first.

We’ve seen that the perception results when the quale interacts with background knowledge: before that happens, there is nothing representational – the quale is size- and color- and temperature--…indeterminate. So there is an extremely *strong* sense in which nothing is given: nothing that even *purports* to give true information about the world exists *without* the co-operation of background knowledge. *A fortiori* there is no *actual* awareness of what is going in the external world without the co-operation of background knowledge. Sensations by themselves are not just epistemically, but *representationally*, inert; they are, as we put it, ‘indeterminate’ until background knowledge sets the relevant parameters. And different bodies of background knowledge can ‘set’ these parameters differently.

All of this is very inconsistent with intentionalism. Intentionalism says that there cannot possibly be a difference in phenomenology unless there is a

difference in representational content. Intentionalism thus assimilates the phenomenological to the representational. Given that there cannot *possibly* (in any possible world) be a difference in phenomenology without one in representational content, it follows phenomenology is realized or constituted by something have representational content. (Otherwise the supervenience of the phenomenology would be inexplicable.) And it also follows that, if *any* bit of phenomenology was non-representational, it would have to be the *only* kind of phenomenal content was non-representational. (For if there were *two* non-representational types of phenomenal content, then there would be a difference in phenomenal content that was, vacuously, *not* matched by a difference in representational content: obviously two bits of phenomenal content that represent *noting* don't represent *different* things.)

But the picture we've sketched contradicts these contentions in a number of way. First, we've seen that phenomenology is *not* typically realized by (let alone identical with) something representational. Perception involves the co-operation of sensations (quale) *and* back-ground knowledge; in such cases, the sensation is not inherently representational, but becomes so only with the help of background knowledge. Thus at least some sensations are not *by themselves* representational. So this contradicts the contention that phenomenal content is always realized by something having representational content.

Also, it would seem that the inherently non-representational qualia that are involved in one perception can be dramatically different from those involved in another perception. So this contradicts the idea that there can be *at most* one kind of non-representational phenomenal content. Thus intentionalism is contradicted on two levels.

To sum up, the intentionalist maintains that a quale *by itself* typically constitutes a representation; for the intentionalist maintains that the thing

having phenomenal content is identical with, or realized by, the thing having representational content. (Thus, for the intentionalist, the thing having phenomenal content cannot possibly fail to be representational– just as the statue cannot fail to be, say, wet if the clay that realizes the statue is wet.) But to say this is fall squarely into the myth of the given: the spurious belief that perceptual representations *do not* implicate background knowledge and sensations *by themselves* are epistemic.

In one place, Sellars himself writes: "Sensations are no more epistemic in character than are trees or tables…"[274] My guess (though the context of that quotation makes it unclear whether this guess is correct or not) is that he had in mind points like the ones we've just stated.

We arrived at these points by doing a great deal of theory. But they are borne out by examples. Under certain lighting conditions, you see a tie – literally *see* it – as being burgundy. Under different lighting conditions, in having a phenomenologically identical perception, you see – literally *see* that tie as maroon (or, in any case, some color other than burgundy). You see a tree as being moderately tall and being (about) twenty feet away. But, given different background knowledge, a phenomenologically identical perception could tell you that you were seeing a gigantic tree that was (about) a hundred feet away.

Suppose that Bob is in his × 1 phase, and that P is a bit of phenomenal content that mediates a perception of a small tree nearby. When Bob is in his ×2 phase, or his ×10 phase, a bit of phenomenal content qualitatively just like P will mediate a perception of a large but distant tree. The *same* bit of phenomenal content was associated with different representational contents. This is inconsistent with intentionalism.

Intentionalists tend to defend their position by seizing on facts like the following. Suppose that in circumstance C1, some sensation is involved in

seeing that a tie is burgundy. And suppose that in C2, some qualitatively *different* sensation is involved in seeing that a tie is burgundy. (This will happen if the lights are dimmed somewhat, or if the over head light is replaced with a less pure, more yellow light: it could happen under any number of circumstances.) Of course, *prima facie* this seems to favor the anti-intentionalist: different representational contents, same phenomenal content. But the intentionalist always responds by saying:

The two perceptions *don't* have the same representational contents. In one case, part of what you are seeing is a light casting an impure yellowish light. In the other case, you may not be seeing that, and you may be seeing a light that is casting a very pure, white light. Or in the one case, the tie may be better *illuminated* than in the other case. So the two qualitatively identical perceptions are representing *different* things: in the one case a *well* illuminated patch of burgundy is represented; in the other, a poorly illuminated patch of burgundy is represented. And *that* is why there is a difference in phenomenal content: for *there is* a difference in representational content.[275]

But this doesn't apply to Bob. Depending on whether he is in his ×1 phase or his ×10 phase, a given a bit of phenomenal content may mediate perceptions with very different representational contents. So there is *sameness* of phenomenal content and a *difference* in representational content. The same bit of phenomenal content is deployed once in seeing a small nearby tree, and another time in seeing distant large tree.

These considerations help us identify the shortcoming in a well known

argument for intentionalism due to (or at least stated by) Alex Byrne. Byrne gives the following argument for intentionalism.

Suppose you are looking at a blue chip. Your seeing the blue of the chip obviously involves your having some kind of phenomenal content. Now imagine a change in the phenomenal content that is involved in your seeing the blue of the chip. If there is such a change, then there is necessarily a change in the color that the chip is represented to as having; it goes from being represented as blue to being represented as (say) green. The same point (mutatis mutandis) applies to *any* perception. If you imagine a change in the phenomenal character of your gustatory perception of the honey, then the honey goes from being represented as sweet to being represented as (say) bitter. One last example: suppose you hear a certain pitch *as* a middle C (you have perfect pitch); if the phenomenal content involved in your hearing that pitch is changed, then *ipso facto* you no longer hear that sound as a middle C but as (say) a middle D. So: no difference in phenomenal content without a different in representational content.[276]

But there are obvious counter-examples to this. You can feel the roundness of the ball and also see it (a bat could hear it, a mole could smell it). But these perceptions have *no* phenomenal overlap, despite considerable (though my no means total) representational overlap. You can *see* distances, and you can also learn distances kinesthetically. Obviously a visual perception has no phenomenal overlap with a series of kinesthetic sensations, despite a certain overlap in representational content.

Also non-occurrent memories have *no* phenomenal overlap with perceptions whose representations they store (or can store, in principle) in

their entirety. Total overlap of representational content: no phenomenal overlap. (Of course intentionalism is defined as the position that two *experiences* cannot differ in phenomenal content without differing in representational content. So it may see that intentionalism doesn't say anything about whether a non-occurrent memory can have the same representational content as, say, a perception without having any phenomenal overlap with it. But this is not the case. Consider a perception P with phenomenal content C and representational content R. Intentionalism says that having P has C wholly *in virtue* of having R. So, for intentionalism, having R is a *sufficient* condition for having P. So, if intentionalism, then nothing *other* than having R is needed to have C. So intentionalism is committed to the view that anything has P as long as it has R. But a non-occurrent memory has R but not P.)

Byrne's argument proves *at most* that *within a single sensory modality* there cannot be phenomenal variation without representational variation. But this thesis is false, it turns out, if we consider different sensory modalities or memories.

But we've seen that Byrne's thesis is false *even* within a single modality. At the same time, it is not an accidental fact, surely, that within a single modality, differences in phenomenal content *tend* to co-vary with differences in representational content. And, admittedly, the most obvious way to explain this is by advocating some kind of intentionalism. So we owe Byrne an non-intentionalist explanation of this co-variation.

What Byrne has shown is not that phenomenal content co-varies *tout court* with representational content, but that phenomenal content co-varies with representational content *all other things being equal*. And I submit that, in this case, the 'other things' consist of one's background knowledge.

Obviously phenomenal content is involved, in an essential way, in our

perceptions; and, obviously, perceptions are involved in an essential way, in our acquiring knowledge of the external world. Now *if* phenomenal contents are to be of assistance in our acquiring empirical knowledge, then presumably they must be paired off in a *consistent* with way their (proximal) causes. If a given kind of sensory disturbance (i.e. a given kind of disturbance of one's retina or epidermis or eardrum…) result now in one kind of phenomenal content, and now in some totally different kind of phenomenal content, then one's phenomenal contents wouldn't be much use in helping us acquire knowledge of the external world. We would be lost. (We can tolerate *occasional* shifts of this kind. But these shifts cannot be the norm: they must occur against a background of stability.) So there must be a reasonably consistent relationship between our phenomenal contents and their (proximal) causes. There must be some rule R that pairs off tissue disturbances (of our sensory organs) with phenomenal contents (and does so in some not ridiculously Byzantine way).

But is there only *one* possible value of R? Do your perceptions have to be associated with their causes in *exactly* the way they are *in fact* associated with them to tell us what they do about the world? Under certain conditions, sound-wave oscillations of 440 Hz cause a certain kind of phenomenal content (call it P1); under those same conditions, sound-wave oscillations of 480 Hz cause a certain kind of phenomenal content (call it P2), and so on. Let R1 be the rule that pairs off phenomenal contents with disturbances of your eardrum in that particular way. Must *R1* be the operative rule of projection? Would a being whose auditory system involved R2 (some rule other than R1) necessarily be mis-hearing everything? It seems that, if R2 is as *consistent* as R1, it can be as good a rule of projection, i.e. it can be involved in the generation of perceptions that are as accurate (if not more so) than those had by a being with R1. In fact, R2 might be better than R1 in this respect. R1

may assign the same phenomenal content to sound-wave oscillations of 440 Hz that it assigns to those of 445 Hz. (Human beings are not even *remotely* capable of registering such minute differences, though other creatures are. So R1 might the rule operative in human hearing.) If R2 is *more* sensitive to differences in sound-wave oscillations – and clearly non-human creatures are better able to hear differences in pitch than we are (and this presumably involves having different *phenomenal* contents in response to different pitches in cases where we would have the *same* phenomenal content in response to both pitches) – and R2 is just as consistent, then it seems extraordinarily arbitrary to say that creatures whose hearing-system uses R2 are just mis-hearing all the time. But that is just what the intentionalist is committed to. Obviously humans hear things the right way, at least some of the time. And obviously our hearing things involves having certain phenomenal contents in response to sound-wave oscillations. The intentionalist says that a creature that has a *different* phenomenal response to the same sound-wave oscillations. The intentionalist *also* says that a difference in phenomenology means a difference in representational content: so those other creatures are *hearing things as being different* and thus – since we are hearing things the right way and they are hearing things differently – as hearing them *wrong*. To clarify this: Suppose a *human's* hearing (under certain circumstances) a middle C involves having phenomenal content P, and that human a C correctly under those circumstances. *For an intentionalist*, a creature that has phenomenal content P* (different from P) under those same circumstances is *not* hearing a middle C; it is hearing some other pitch. So that creature is hearing the wrong thing. Thus if that human's audition involves R1, then every other right-hearing creature's hearing must involve R1 as well. No other correspondence rule, no matter how fine-grained (even if it is much more nuanced than ours) or consistent will do.

But again this seems deeply arbitrary. If another creature's phenomenal responses to sound-wave oscillations are just as consistent as ours and just as fine-grained, then by any reasonable conception of information, it can acquire as much information about the world through hearing as we can; and it then seems arbitrary to say that it systematically mis-hears just because it is employing R2 and not R1. And if a creature's hearing involves a *more* fine-grained, and more internally consistent rule than ours, then by any reasonable conception of information, it can acquire *more* information (through hearing) than we can about the world; and it then becomes all the more arbitrary to say that such a creatures is always hearing things the wrong way just because it uses R3 not R1. And there is no doubt that there are, in principle, rules *other* than the one operative in human hearing that *are* more fine-grained than ours and as consistent than ours. (to say that a rule R is more 'fine grained' than a rule R* is to say that, for some n, if sound-waves differ from each by less than n, but still differ, R will, whereas R* will not, pair off those sounds with different phenomenal contents. There can be no doubt that, by this definition, there are creatures whose hearing involves much more fine-grained rules of correspondence than ours – bats, dogs, rabbits.) So there seems no reason in principle to *deny* that creatures that hear *better* than us might have different phenomenal contents from us in response to a given sound-wave under a given circumstance. But intentionalism says *no difference in phenomenal content without a difference in representational content.* And the motivation for this view is that the phenomenal reduces, after a fashion, to the representational. But apparently it does not. [277]

I would suggest that what Byrne has shown is that *as long as* a given rule of projection is operative, then the phenomenal 'tracks' the representational, i.e. if perceptions have different phenomenal content, then they have different representational contents. This is to be expected since an

inconsistent rule of projection – one that assigned *different* phenomenal contents to the *same* sound (heard under the same circumstances) – would not be of much use (in learning about actual sound-phenomena). But what Byrne doesn't register is that *if* the rule of projection is changed, then the representational value of a given phenomenal content can change. Suppose that Bob is a human being who has (by human standards) good hearing. Let R1 be the operative rule projection. Suppose that neurosurgeons change Bob's auditory system. Now the rule of projection that is operative in his hearing is the one operative in a bat's. Obviously Bob's hearing a lot better than it used to be. But given what we said a moment ago, it seems highly likely that, in this new and improved version of Bob, a phenomenal content that *used* to be involved in hearing (say) a middle C is *now* involved in hearing a middle D. So when the *rule of projection* is changed, Byrne's thesis no longer applies: phenomenology tracks differences in representational content only *given* a certain rule of projection. In any case, this seems plausible.

We've said a lot about the role of 'rules of projection' in the assigning of representational contents to phenomenal contents. What we've said on this matter is consistent with Sellars' attack on the myth of the given. Sellars' attack, we've seen, comes in two strengths. Weak version: it is not typically true that qualia, by themselves, constitute *awarenesses* (knowledge) of anything external – something extra is needed to make a quale be *epistemic*. Strong version: it is not typically true that qualia, by themselves, constitute *representations* – something extra is needed to give them representational content.

We've seen reason to hold that both versions are correct. But we seem to have a problem. We've apparently given two different answers to the question 'what is that extra something which makes a non-representational

(or non-epistemic) quale be representational?' One answer we gave is: *background knowledge*. The other answer is: *a rule of projection.*

To deal with this, we need only be clear on what we've said so far. The rule of projection doesn't assign *representational contents* to disturbances of our sensory surfaces. The rule of projection assigns only *phenomenal contents* – qualia – to such disturbances. It is *background knowledge* that assigns representational contents those qualia. So the 'rule of projections' that we've been talking about have an altogether different function from background knowledge, though of course the two are part of a co-operative enterprise.

Conclusion

Intentionalism says that, typically, qualia *by themselves* have determinate representational content: background knowledge isn't needed to *endow* qualia with representational content. And, of course, in so far as qualia have representational contents *ab initio,* background knowledge cannot affect what representational content they have. For *if* background knowledge *did* have a hand in endowing a quale with representational content, it could only be by way of some openness, some representational indeterminacy, on the part of the quale; and, by definition, a quale with a 'fixed' representational content has no such indeterminacy or openness. Thus the picture given to us by the intentionalist is that, typically, qualia don't *need* background knowledge to fix what they represent, and they won't *let* background knowledge fix what they represent. Qualia thus become *foundational*, performing their representational and epistemic roles in splendid isolation of any epistemic context. But it is just this kind of foundationalism that Sellars (and, after him, many others) have shown not to be tenable. Intentionalism is a strong form of

foundationalism; it is guilty of the myth of the given, in both its forms. Therefore *either* intentionalism is to be rejected *or* the anti-foundationalism of Sellars and others – now so widely held – must be false.

Burge, Tyler. 1993. “Content Preservation”. *The Philosophical Review*, Vol. 102, No. 4, (Oct., 1993) 457-488.

Byrne, Alex . 2001. ‘Intentionalism Defended’. *The Philosophical Review*. Vol. 110, No. 2 . (April 2001).

Sellars, Wilfred. 1963a. “Empiricism and the Philosophy of Mind”. In *Science, Perception and Reality.* New York: Routledge and Kegan Paul.
-- 1963b. “Some Reflections on Language Games”. In *Science, Perception and Reality.* New York: Routledge and Kegan Paul.

Propositions, Logical Truth, and the Analytic-Synthetic Distinction

Abstract

It is hotly debated whether or not there are non-spatiotemporal entities. It is not debated whether the so-called linguistic theory of logical truth is correct: everybody (nearly enough) agrees that it is false. This paper shows that one must *either* accept the existence of propositions (conceived as platonic entities of some kind) *or* one must accept the linguistic theory of logical truth. So propositions, and therewith a realm of platonic entities, must be countenanced. Given this last fact, Quine's argument against the analytic-synthetic distinction proves to involve a fatal non-sequitur. Further, given that there are propositions, it becomes possible to give a viable analysis of "analyticity".

I. The sentence "snow is white" is meaningful. But is there some *object* that is its meaning? We know tables and chairs exist. But do the *meanings* of "snow is white" and "grass is green" exist alongside them?

Given *only* that sentences are meaningful, it doesn't follow that there such things as meanings. An analogy might help. For a work of art to be beautiful isn't for it to be associated with some *object*. It isn't as though the work of art is one thing, its beauty another, and the work of art is beautiful because it is somehow associated with this other entity. One might say that the property of being meaningful is like that of being beautiful: for a sentence to be meaningful is no more for it to be associated with some *object* than a painting's being beautiful consists in *its* being associated with some object.

[278]

The meanings of sentences, if such things exist, are referred to as "propositions". So if in fact there is some object that is the meaning of "snow is white", that object is a "proposition".

Many philosophers deny the existence of propositions. In their view, it is as absurd to suppose that there is some *object* that is the meaning of "snow is white" as it is to suppose that there is some *object* that is the beauty of the Mona Lisa.

The purpose of this note is to show that propositions *do* exist. I will show that, if one denies the existence of propositions, one is committed to a doctrine that is now (rightly) believed false: the so-called "linguistic theory of logical necessity."

If *anything* is universally agreed on among philosophers, it is that the linguistic theory is false. What is *not* agreed on is whether there are such things as propositions: that issue is maximally controversial. This paper will show that, if one denies the existence of propositions, one has no choice but to advocate the linguistic theory. So this paper will try to tackle a controversial issue (whether there are propositions) by attempting to draw a tight connection between it and a non-controversial issue (whether the linguistic theory is correct).

If I am not mistaken, we will find that Quine's celebrated argument against the analytic-synthetic distinction is a failure, as it is implicitly committed to the linguistic theory of logical necessity.[279] Thus the present paper, if cogent, will rehabilitate the analytic-synthetic distinction – or, at least, will undermine one of the more prominent arguments *against* that distinction.

II. The sentence "triangles have three sides" is in a different category from the sentence "Smith is bald." It would be incoherent to deny the former; it

would be coherent to deny the latter. Consequently, what the former says *must* be the case, and what the latter says need not be the case.

It is through sense-perception that we know Smith to be bald. We *see* that Smith is bald or we sense-perceive evidence of his baldness. But how do we know of the truth of what is said by "triangles have three sides"? We obviously don't know it by inspecting triangular objects. It isn't through empirical research that we know it.

Rationalists say: it is through "pure reason" – through some non-sensory cognitive faculty – that we know that triangles have three sides. Empiricists cannot take this position. For empiricism is, by definition, the view that all knowledge is acquired through the senses.

One way empiricists have dealt with necessary truths is by turning them into *linguistic* truths. Sentences like "triangles have three sides" are plausibly seen as being true "by definition". If one were to say "triangles do not have three sides", one would be "sinning against language", as Ayer put it [280]; one would be using words in deviant ways. So "triangles have three sides" is made true by facts about language – facts about how people use words. It is true wholly in virtue of "linguistic convention".

III. This theory is now (rightly) rejected. Later on we will discuss why it is wrong. What I'd like to do now is to show that, unless one believes in propositions – unless one believes that, in addition to tables and chairs and iguanas, there are propositions – one must advocate this false theory.

Let us refer to people who believe in propositions as "propositionalists", and to people who do not as "anti-propositionalists".

Once again consider the sentence:

(T) "Triangles have three sides."

A propositionalist says:

(T) is necessarily true because it has for its meaning a necessarily true proposition, namely: *triangles have three sides.* That proposition is made true by facts about mind-independent concepts. The concepts *triangle*, and *three* and *side* are so interrelated that (T) must come out true, given that it bears the proposition *triangles have three sides*. So language makes (T) be true only in the sense that it associates that sentence with a necessarily true proposition. That proposition is true independently of anything having to do with language.

An anti-propositionalist cannot say that. He cannot say that what makes (T) be necessary is that it is associated with a necessarily true proposition. For the anti-propositionalist doesn't believe in propositions at all. So the anti-propositionalist has to figure out some other way of answering the question: What makes (T) be true?

What options does the anti-propositionalist have? He cannot say that (T) is made true by facts about the interrelations of platonic entities – of properties or concepts (in the non-psychological sense). Why can he not say this? To take this line would be to grant the existence of propositions. Presumably, the proposition *triangles have three sides*, if it exists, just *is* some structure consisting of the concepts (or properties) *triangle*, *side*, and so forth; so it just *is* these objects, interrelated in a certain way. So if an anti-propositionalist says that interrelations among platonic entities are what make (T) be true, then he is conceding that propositions *do* exist, and he is no longer an anti-propositionalist.

Surely *triangles have three sides*, if it exists at all, is built up out of the concepts *triangle*, *three*, *side* and so on. The propositions *triangles have three sides* and *Jim's pool has the shape of a triangle* – if they exist at all -- have something in common, corresponding to the fact that *triangle* is a constituent of both of them. So I don't see how one could deny that *triangles have three sides* consists of the concepts (or properties) *triangle*, *three*, and *side*. And it is truistic that, if that proposition exists at all, these objects are arranged in a certain way in it.

So what else is there, besides platonic entities and the relations that hold among them? There is also the spatio-temporal world. (And for reasons we've just noted, anti-propositionalists tend to say that the spatio-temporal world is *all* there is.) So the anti-propositionalist must look to facts about that domain, in his search for the truth-maker of (T). Trivially, there are two kinds of spatio-temporal facts: those that have to do with *language* – facts having to do with what expressions mean and with how people use them – and those that do not. Let us see if the anti-propositionalist can find what he needs in either of these domains.

Suppose the anti-propositionalist says that *non*-linguistic spatio-temporal facts are responsible for the truth of (T). To say that non-linguistic empirical facts are what make (T) be true is to say that (T) is in the same category as (B) "Bachelors are emotionally unfilled". To say that non-linguistic spatio-temporal facts are what make (T) be true is to say, or at least imply, that to know that (T) is true, one has to *examine* the non-linguistic portion of the spatio-temporal world. And if one has to examine that portion of the world to know that (T) is true, then it is contingent (more exactly, it is *epistemically* contingent) that (T) should be false. For if one *must* study the non-linguistic part of the spatio-temporal world to know the truth of (T), then it is by all

means possible to coherently *conceive* of (T)'s turning out false; for the negation of anything that can *only* be known through empirical inquiry is coherently conceivable, practically by definition. But everyone admits that (T) is not contingent, epistemically or otherwise; (T) is not – and no one (with the *possible* exception of Mill) denies this - in the same category as "bachelors are emotionally unfulfilled".

So if spatio-temporal facts are what make (T) be true, the facts in question cannot be *non*-linguistic facts. So they must be *linguistic* facts. So (T) is true because of facts about *language*. There are right and wrong ways to use expressions. Ultimately, whether it is right or wrong to use an expression in a certain way is a matter of convention. (I say "ultimately" because, although there are no conventions *directly* governing how to use *complex* expressions like "penguins would be able to fly if they had bigger wings", there *are* conventions that directly determining how to use the corresponding simple expressions: "penguin", "fly" and so on.) If you say "triangles do not have three sides" or "triangles have two sides", you "contravene" these conventions, as Ayer put it. Your statement is counter-linguistic and ungrammatical.

In a word, if one rejects propositions, one must embrace the linguistic theory of necessity.

IV. The linguistic theory is not without merit. One doesn't need any *empirical* knowledge, apart from knowledge of semantic facts (linguistic conventions, semantic rules), to know that (T) is true and that the sentence "triangles have four sides" is false. The linguistic theory seems to account for this fact. So that theory escapes the absurd doctrine (sometimes attributed to Mill) that (T) is in the same category as "bachelors are emotionally unfulfilled." Further the linguistic theory seems to show how a strictly

empiricist epistemology can account for our *knowledge* of necessary truths.

But if anything is known with certainty in philosophy, it is that the linguistic theory is false. Its faults are extremely well-documented.[281] Nonetheless it might be appropriate to give a few of the many cogent arguments that have been directed against it. I will also put forth an argument of my own. My argument isn't as air-tight as some of the existing ones, but it is not wholly without force.

A famous argument is due to Quine.[282] There are infinitely many logically true sentences. Convention can be responsible for, at most, a finite number of these. We haven't existed long enough to have enacted infinitely many conventions. (We couldn't exist long enough for that.) So the rest are not *directly* made true by convention. They are, at most, *indirectly* made true by convention. So they *follow* from conventions. But to say that they *follow* from conventions is to say that conventions *plus* logic are what make them true. So conventionalism reduces logic to conventions *plus* logic, and thus doesn't reduce logic to conventions at all.

Butcharov has a clever argument against the linguistic theory.[283] (T) could have meant that penguins are good at chess. After all, semantic facts are obviously quite contingent. So (T) is necessary *only if* it means that triangles have three sides. So what is necessary is not (T); what is necessary is the statement:

(*) *If* (T) means that triangles have three sides, *then* (T) is necessary.

But (*) cannot be made true by linguistic facts. To deny this is to incur a vicious regress. For obviously the *sentence*

(**) “If (T) means that triangles have three sides, *then* (T) is necessary”

could have meant that aardvarks are good at miniature golf or that the moon is made of cheese. So that *sentence* is not necessarily true. What is necessarily true is:

(***) If (**) means (*), then (**) is necessary.

If we say that (***) is made true by linguistic convention, then we are embarking on a regress. If we *deny* that (***) is made true by linguistic convention, we are rejecting the linguistic theory of necessity.

Another argument against the linguistic theory is this.[284] According that theory, sentences like “triangles do not have three sides” contravene the rules of language. Now a string of verbiage that contravenes the rules of language is *ipso facto* not well-formed and is therefore nonsense. But “triangles do not have three sides” is not nonsense; for it is false, and what is false is not nonsense. “Bob triangles gladly the” is nonsense. “Triangles do not have three sides” is not nonsense. No *linguistic* fact makes “triangles do not have three sides” be false, and no linguistic fact makes “triangles have three sides” be true. To break a convention associated with language L is to produce something that is not a sentence of L. “Triangles do not have three sides” is a sentence of English: a false one. So that sentence breaks no conventions of English; therefore it is not false *by reason* of its breaking any conventions of English; therefore its negation (“triangles have three sides”) is not true *by reason* of its conformity to any conventions of English.

There is a more fundamental reason why the linguistic theory fails. Triangles had three sides before there was language, and they will have three sides after there is language. And this obviously isn’t a contingent fact. So

language cannot be responsible for this necessity.

I would like to give an original argument against the linguistic theory of necessary truth.[285] I submit that, embedded within the concept of language, is the notion of necessity. Consequently to say that "language" underlies necessary truth is to speak in a circle.

I warn that this argument is not as cogent as the others. This is because it presupposes a certain conception of what a language is.

What is a language? One answer is this: a set of rules that (inter alia) associate expressions with truth-conditions: rules like

"…Socrates…" is true iff Socrates has the property…x…

and

"x is red" is true iff the object denoted by "x" is red.

(So "Socrates smokes" is true iff Socrates has the property of smoking. And "Jim's wig is red" iff the object denoted by "Jim's wig" is red.)

Any human language must be learnable. Any learnable language contains only finitely many primitive rules. So any human language must contain finitely many primitive rules. But any human language contains infinitely many sentences that are assigned truth-conditions by those rules. Those rules cannot *directly* assign truth-conditions to all of those sentences. So the fact that those sentences have those truth-conditions *follows* from those rules, while not being *directly* established by them. In other words, those rules assign those truth-conditions to those sentences *via* principles of logic, and thus *via* necessities. So the concept of language – at least a language that, like English or Albanian, contains infinitely many sentences and finitely many

primitive rules – is to be defined in terms of the concept of necessity. Thus one cannot non-circularly attempt to trace necessities to language.

It must be said that the very people who propounded the linguistic theory of necessity – Ayer, Carnap, Schlick – *did* think of language in exactly the way I have just said. They thought of language as a formal calculus. Primitive expressions are assigned meanings by simple rules, and complex expressions are assigned meaning by way of *derived* rules or *theorems* of that calculus. Thus my argument shows that the linguistic theory fails on its own terms or, at least, on the terms of those who propounded it. So to that extent, the argument just given is not question-begging.

V. This discussion makes us aware of an interesting irony the history of analytic philosophy. Quine, as we noted a moment ago, gave a compelling argument *against* the linguistic theory of necessary truth. But a few years later, in a publication that would come to be considered one of the two greatest works 20th century philosophy[286], Quine *rejects* the idea that there are propositions. He says that the notion of "meaning" (and *a fortiori* the notion of sentence-meaning or propositionality) is incorrigibly "obscure" and cannot pass methodological muster. Having rejected the notion of meaning, Quine then attempts to prove that there is no principled way to distinguish between analytic and synthetic sentences.

His argument is basically this. Consider the sentence "all bachelors are unmarried". Supposedly this sentence is analytic. If it is analytic, what makes it so? The obvious answer is this: the *meaning* of "bachelor" is related in a certain way to the *meaning* of "unmarried". "Unmarried" means *unmarried* and part of what "bachelor" means is *unmarried*. So "all bachelors are unmarried" is synonymous with some sentence of the form "anything that is an F and a G is a F", and obviously such a sentence is logically true, and thus

a necessary truth.

But Quine says that, because the notions of "meaning" and "proposition" are so bankrupt, we cannot countenance this analysis. Thus we must look for some other way to ground the analyticity of "all bachelors are unmarried". The only possibility is that "bachelor" is *synonymous* with some expression of the form "…unmarried…" But it turns out that for two expressions to be synonymous is for them to be intersubstitutable *salva analycititate.* In other words, E and E* are synonymous only if, given an analytic sentence of the form "…E…", replacing E with E* results in an analytic sentence.

So we end up defining analyticity in terms of synonymy and synonymy in terms of analyticity. This definition will not do, as it is viciously circular. We must conclude that analyticity is not non-circularly definable and is therefore a bankrupt concept.

We've seen that if one *rejects* the notion of a proposition, one is committed to the *linguistic* theory of necessary truth. And Quine himself showed that theory to be untenable. So Quine has hoisted himself up by his own petard.

As we have seen, given that the linguistic theory is false, it follows that there are propositions and thus meanings. Once this is granted, then Quine's argument against the analytic-synthetic distinction is blocked. Why is "bachelors are unmarried" analytic? Because the *meaning* of "bachelor" is related in a certain way to the *meaning* of "unmarried": the concept expressed by "bachelor" is such that, if x falls under that concept, then as a matter of epistemic necessity, x must fall under the concept expressed by "unmarried." Quine rejects this answer because he rejects the notion of meaning. But we've seen that this rejection commits him to the linguistic theory of truth. So the notion of meaning must *not* be rejected.

There is another point to make about Quine's attack on the analytic-synthetic distinction. In *Two Dogmas,* Quine concedes that formally true

sentences (what he calls “truths of logic”) are analytic. He concedes that “anything that is both an F and a G is an F” is analytic. (Obviously if Quine denied this, he wouldn’t have many followers.) What Quine denies is that any sentence that is *not* formally true can be analytic. “Bachelors are unmarried” has the same form as “bachelors are grumpy”. So, for Quine, “bachelors are unmarried” is not analytic.

In saying that “anything that is an F and G is an F” is logically true, Quine is making a very modest concession. But he cannot make this concession without undermining the point he is arguing for (in *Two Dogmas*). We can show this by using Quine’s own argument against conventionalism against *him.*

There are infinitely many formally true sentences. (Quine concedes this.) Non-linguistic spatio-temporal facts cannot be what is responsible for the truth of a *formally* true sentence. A *formally* true sentence is, by definition, one whose truth is guaranteed by its form and whose truth is, therefore, not a function of any non-linguistic spatio-temporal entity. Obviously Quine cannot say that such sentences are true in virtue of what *meanings* (propositions) they are associated with. Since non-linguistic spatio-temporal fact don’t make them true, and since meanings (and thus facts about platonic entities: for as we’ve seen, meanings/propositions just *are* structures consisting of such entities) don’t make them true, the only thing left to make them is true is *language*: facts about semantics and word-usage and the like. But such facts – to echo Quine’s argument against conventionalism – can be *directly* responsible for the truth of only finitely many formally true sentences. But there are infinitely many such sentences. Let S be the infinitely large set of formally true sentences for whose truth linguistic facts are not *directly* responsible. So such facts are, at most, indirectly responsible for the truth of the members of S. The truth of those sentences is not *directly*

guaranteed by linguistic facts, but *follows* from such facts. This means that such facts, *plus* truths of logic, are what make the members of S be true. But, on pain of incoherence, those truths of logic cannot themselves be seen as holding in virtue of language. Nor can they be seen as holding in virtue of non-linguistic spatio-temporal facts. (For if those truths of logic held in virtue of non-linguistic spatio-temporal facts, then the truth of the members of S true would itself turn on matters of spatio-temporal fact. But it doesn't, as we've seen.) So those truths of logic cannot be regarded as holding in virtue of linguistic spatio-temporal fact *or* in virtue of non-linguistic spatio-temporal fact. They can only be seen, then, as holding in virtue of facts of a *non*-spatio-temporal kind: facts about *platonic* entities and their interrelations. As we saw earlier, once we concede the existence of such facts, we have in effect conceded the existence of propositions.

I will end by considering an objection to this line of thought:

A *formally* true proposition isn't made true by *anything* other than its form. So it isn't made true by (a) non-linguistic spatio-temporal matters of fact, or (b) linguistic matters of spatio-temporal fact, or (c) facts about platonic entities. The argument you just gave assumes that formally true propositions are made true by one of those three things, and that assumption is false.

That assumption is *not* false, and this becomes clear when we examine the concept of sentential *form*. When you say that a sentence is true in virtue of its "form", that obviously doesn't mean that it is true in virtue of its *shape.* Nothing would be easier to imagine a language in which the following sentence token "anything that is an F and a G is an F" meant *grass is orange.* And, of course, in such a language "anything that is an F and a G is an F" wouldn't be true at all, let alone "formally" true. Basically, "form" doesn't

mean “morphology”, at least not in this context.

When it is said that “anything that is green and bald is green” is “formally” true, that is an extremely loose way of saying that the only terms in that sentence whose *meanings* are relevant to its truth are the so-called “logical constants”: “anything”, “is”, “and”, and so on. Why *those* particular terms are held to be form-defining and not, say, “green” or “bald” is a matter of considerable debate.[287] But the basic idea is that, no matter what terms we substituted for “bald” and “green”[288], the resulting sentence would be true. Thus “bald” and “green” figure only vacuously in that sentence. But, of course, the “logical constants” cannot typically be replaced without loss of truth; replacing the “and” with an “or”, for example, would result in a falsity.

The basic idea is that what we (misleadingly) describe as “formal” truth is actually a *certain* kind of truth in virtue of meaning. “Anything that is green and bald is green” is “formally” true because of facts about the *meanings* of certain of its terms – “and”, “anything”, and so on – guarantee its truth. Thus what makes that sentence be true are certain facts about its meaning. This, of course, is consistent with our point that logical truths hold in virtue of facts about *platonic* entities and their interrelations.

VI. I have argued that, if one denies the existence of propositions, one is committed to the linguistic theory of logical truth (a theory everyone agrees is false). An important part of my argument was the following statement:

(*) [T]he proposition *triangles have three sides*, if it exists, just *is* some structure consisting of the concepts (or properties) *triangle*, *side*, and so forth; so it just *is* these objects, interrelated in a certain way. So if an anti-propositionalist says that interrelations among platonic entities are what make

(T) be true, then he is conceding that propositions *do* exist, and he is no longer an anti-propositionalist.

Obviously (*) is a reasonable enough point. But it has been denied. So to make my argument cogent, I must give a defense of (*).

First of all, on what grounds has (*) been denied? Some have held that propositions are sets of worlds or are functions from worlds to truth-values: the so-called possible-worlds conception of propositions. And an advocate of that view would, indeed, *not* hold that the proposition *triangles* have three sides consists of the concepts *triangles* and *three* and so on.

But that view has notorious problems. The main problem, of course, is that it doesn't distinguish between analytically equivalent propositions. (So, on the possible worlds approach, *2+2=4* and *triangles have less than seven sides* are the same proposition. For they are true in exactly the same worlds: all of them.) Another is that it is circular if it construes worlds as sets of propositions. For then a proposition is analyzed as a set of sets of truth-values or as a function from sets of propositions to truth-value: both analyses are blatantly circular. On the other hand, the possible worlds approach is false, or at least radically implausible, if it construes propositions as sets of *concrete* worlds. For it is very hard to see how we would know anything about concrete objects lying outside this world.[289] And, of course, we must be suspicious of semantic theories that require us to posit entities for which there is no independent corroboration.

But we've just encountered yet another problem with that view. It is clear that *triangles have three sides* and *squares have four sides* and *Jim's pool has three sides* all have something in common, corresponding to the fact that *side* is a constituent of all of them. But the possible-worlds analysis of propositions strips propositions of any constituent-structure or, at any rate, of

the kind of constituent structure that would guarantee – what we know to be true – that *side* is a constituent of all of these propositions. For the possible worlds view says: all of these propositions are just sets of worlds. But the concept *side* is not in any clear sense a constituent of a set of worlds. So our analysis of propositions – that they are structures consisting of properties (or concepts) and objects (that *triangles have three sides* consists of the concepts/properties *triangle*, *side* and so on, interrelated in a certain way) – doesn't beg questions against the possible worlds approach but, on the contrary, provides an additional *independent* reason to reject that approach.

A related problem with the possible worlds approach is that it isn't consistent with the compositionality of thought. Obviously one's grasping (a) *triangles have three sides* is *cognitively* related to one's grasping (b) *squares have four sides* and also to (c) *Jim's pool is triangle-shaped.* One's grasp of the concept *side* is deployed in one's grasping *both* (a) and (b), and one's grasp of the concept *triangle* is grasped *both* in one's grasping (b) and (c). Now unless both (a) and (b) had as a constituent the concept/property *side*, and unless both (b) and (c) had as a constituent the concept *triangle*, these facts would be inexplicable. The fact that a grasp of a single concept or property is used in grasping *different* propositions suggests that propositions are built up out of concepts or properties. And this fact is completely inexplicable on a possible worlds view. So we are vindicated, I think, in holding that propositions – if they exist – just *are* structures consisting of properties (concepts). Thus, the anti-propositionalist cannot say that interrelations among concepts/properties are what make "triangles have three sides", and other sentences of its kind, be true.

Most importantly, practically no anti-propositionalists would *want* to say that "triangles have three sides" is made true by facts about platonic entities. For belief in such entities is very much not in the spirit of anti-

propositionalism. Typically, one is an anti-propositionalist precisely *because* one doesn't countenance platonic entities. Propositions come under suspicion because they are typically seen as being platonic entities.

To sum up, the anti-propositionalist is barred from saying that facts about platonic entities make (T) be true. And the anti-propositionalist doesn't even *wish* to take that view.

References

Ayer, Alfred Jules. 1952 Dover Publications New York

Blackburn, Simon. 1984. *Spreading the Word.* Oxford: Clarendon.

Boghossian, Paul Artin. 1997. "Analyticity". In *A Companion to the Philosophy of Language.* Bob Hale and Crispin Wright (eds.). Oxford: Blackwell.

Bonjour, Laurence. 1998. *In Defense of Pure Reason.* Cambridge: Cambridge University Press.

Dummett, Michael. 1978. "The Significance of Quine's Indeterminacy Thesis". In *Truth and Other Enigmas.* Cambridge: Harvard University Press.

Pap, Arthur. 1958. *Semantics and Necessary Truth.* New Haven: Yale University Press.

Peacocke, Christopher. 1999. *Being Known.* Oxford: Clarendon Press.

Quine, Willard van Orman. 1935/1966. "Truth by Convention" in *The Ways of Paradox and Other Essays.* Cambridge: Harvard:.
--1951/1990. "Two Dogmas of Empiricism". In *The Philosophy of Language* (A.P. Martinich ed.) Oxford: Oxford University Press.

Wittgenstein, Ludwig.1953. *The Philosophical Investigations.* Translated by G.E.M. Anscombe. Oxford: Basil Blackwell.
-- 1958. *The Blue and the Brown Books.* New York: Harper and Brothers.

Causal Efficacy and the Erroneousness of Functionalism and Content-externalism

I. There are some compelling *prima facie* reasons to think that functionalism and certain forms of externalism strip mental content of any causal powers. If that is the case, then we have a powerful to reject those doctrines. Jackson and Pettit have produced a brilliant argument purporting to show that, in fact, functionalism and the relevant forms of externalism *are* compatible with the presumption that mental content is causally efficacious. Their argument involves an original and (I believe) correct point about causation. But I believe that their *application* of this correct point to this particular set of problems is flawed: they have not, I will argue, given us any reason to think functional and the relevant forms of externalism *are* in fact compatible with the presumption that mental content has causal powers.

II. It seems clear that, in some cases, brain-states realize mental contents. (I wish the word "state" to be taken as broadly as possible – to apply to both enduring brain-structures and occurrent phenomena.) My belief that 1+1=2 is realized by certain brain-states. My belief that water is wet is realized by a different set of -brainstates.

Let x be an arbitrary brain-state According to functionalism, for x to realize mental content C is for x to have a certain causal role. For x to have a certain causal role is, at least in part, for x to have such and such effects. So x realizes a belief that bread nourishes at least partly in virtue of the effects that x has – for example, it leads to my choosing to eat bread (in certain contexts).

There is an obvious *prima facie* problem with functionalism. Mental content is causally efficacious. In any case, this is what we are inclined to think. My believing that bread nourishes is causally effective: it leads me to

drive to the story and buy bread, to eat said bread, and so on. Given this, suppose that brain-state x has effects E1…En. x's having C presumably has a *causal* hand in bringing about at least some of E1…En. As we just saw, supposing that x is a brain-state realizing your belief that bread nourishes, x's being such a belief is what *causes* you to buy bread, eat it, and so forth. But if x's having those effects is what x's realizing that content *consists in*, then x's realizing that content cannot possibly be what *causes* E1…En to come about. If x's realizing a belief that bread nourishes *consists* in its causing me to drive to the store and buy bread, and to eat bread, and so on, then x's realizing that belief cannot possibly be what *causes* me to drive to the story, to eat the just purchased bread, and so on. So if functionalism is right, my belief that bread nourishes ends up *not* having a causal hand in my going the store to buy bread; and, by an exactly similar argument, it ends up *not* having a causal hand in my wanting to eat bread, or bake bread, and so on. It ends up not having a causal hand in any of the contexts where, quite obviously, it *does* have a causal hand.

The nature of this apparent problem for functionalism becomes clear if we are careful to make a certain distinction. Strictly speaking, it is not *things* that have causal powers. It is not *the rock* which breaks the window, but the rock's moving with a certain velocity, and having a certain mass, and so on. *The rock* doesn't do anything. What does the work is really the rock's having certain properties – a certain mass, solidity, velocity, and so on – along with various other things having various properties (e.g. the window's having a certain micro-structure).

Not every one of the rock's properties is causally effective in *every* causal transaction involving the rock. Suppose the rock has a certain dimple on one of its sides. In some contexts, the rock's having that property will be causally effective. (For example, if you pour water down that side of the rock, the

presence of the dimple will affect the course of the water.) But in other contexts, the presence of that dimple will *not* be causally effective. (For example, if you drop the rock in a vacuum, the presence of the dimple won't have any effect on the trajectory of the rock.)

Suppose, once again, that x is a brain-state realizing my belief that bread nourishes. Strictly speaking, x doesn't do anything: what does the work is x's having certain properties – e.g. x's having certain electrical, chemical, and cellular properties. Not every causal transaction involving x is one where x's *electrical* properties are causally effective. Perhaps in certain contexts it is x's mass, not its electrical properties, that are causally relevant. Similarly, not every causal transaction involving x is one where x's *mental* properties are causally effective. In some contexts, x's mental properties might be causally irrelevant: what is doing the work in those contexts might be x's having a certain electrical charge or its having a certain specific shape. So there is a difference between *x's* being involved in some causal transaction, on the one hand, and *x's realizing a belief about bread* being causally effective in that context. The latter might or might not be causally effective in that context.

But, of course, we want to believe that, in at least some cases, my believing that bread nourishes is causally effective. So, in effect, we want to believe that x's *having a certain mental content* is specifically what is causally effective in certain contexts. What leads me to eat the bread in front of me is my believing that bread nourishes. Here what is effective is not just x, but is x's having the right mental properties.

Here is where the problems for functionalism begin. The functionalist says that x's realizing that belief *consist* in its having certain effects: it *consists* in, for example, its leading me to eat the bread in front of me (instead of putting it in a lead box). But if x's realizing that belief *consists* in x's having those effects, then x's having that content cannot *cause* those things to happen – it

cannot be what *causes* me to eat the bread.

Let us sum up. Functionalism says that for x to such and such mental content is for x to have such and such a causal role – it is, at least in part, for x to have such and such effects, to bring about such and such states of affairs. If that is right, then x's having that content cannot be what *causes* those states of affairs to come about. x's being a belief cannot be what *causes* me to choose to eat the bread; for x's being such a belief *consists* in its causing me to make that a choice. So x's being such a belief cannot be what *causes* me to make that choice. By an exactly argument we can show, for *any* mental state x* realizing content C* that, if functionalism is right, x*'s realizing C* becomes causally inert.

So it seems that, if functionalism is right, then mental content is *inert.* If content is *identified* with causal role, then content cannot be what *brings about* causal role. But in that case, content doesn't *do* anything; it becomes inert, a free-wheel or epiphenomenon. But plainly content is *not* inert; plainly content does a lot. So functionalism must be wrong.

III. Another doctrine – content externalism – also seems to make content inert. Smith and Twin-Smith are atom for atom duplicates. But Smith is thinking about Mary, while twin-Smith is thinking about twin-Mary. The reason is that Smith's brain-state has one distal cause (Mary), while twin-Smith's has another (twin-Mary). The difference in distal cause makes for a difference in content. But this difference in content seems to be inert. What does all the work is *local*: the spatially and temporally local property of Smith's brain are what cause him to pick up the phone and dial those numbers; the same is true (mutatis mutandis) of Twin-Smith. The differences in distal causes are irrelevant: everything that does causal work is in the here

and now – is lodged inside Smith's and twin-Smith's crania.

But if the externalist is right, Smith and twin-Smith have different mental *contents*, corresponding to the differences in distal causes. So, if externalism is right, that content is inert: the fact that Smith is thinking about Mary, as opposed to twin-Mary, about water as opposed twin-water, is inert. What is *not* inert – what does all the causal work – is precisely what Smith and twin-Smith have in common; and what they have in common is neutral between Mary and twin-Mary. So to the extent that mental content is to be understood in externalist terms, mental content doesn't do anything.

IV. Jackson and Pettit deal with these problems in a brilliant way. We must distinguish causal *efficacy* from causal *relevance.*

In everyday discourse, we say things like the following.

"The liquid in the glass froze (thus expanding); and that is what is caused the glass to shatter."

The truth is: what caused the glass to shatter is something much more specific; it is the fact that, in certain very specific places, certain very specific pressures were put on the walls of the container.

Consider the statement:

"The presence of a flame under the container caused the liquid to boil."

Well, what really caused the container to heat up is much more specific: certain specific displacements of energy led to some other specific displacements of energy; and so on. Not every displacement of energy constitutive of the flame caused the increase in temperature. It wasn't really

the flame which did the work: it was certain very specific micro-events constitutive of the flame.

One last example:

"The pianist hit a wrong note because *somebody* coughed too loud."

Well, what *really* caused the pianist to hit a wrong note is much more specific. *Smith* coughed too loud – it wasn't just *somebody*

Causes are much more specific than our statements typically indicate; and the things we describe as "causes" – *somebody's* coughing too loud, a *flame's* being underneath the container, the freezing of the liquid in the glass – are not really what is doing the causal work; there is always something much more specific going on; some specific molecule (or atom or quark…) bouncing off of some other molecule (or atom…). In fact, in the end, the only things that are doing causal work are specific, minute displacements of mass-energy. *The flame* is much too expansive and gross an entity to be responsible for the fact that this, that, or any other specific part of the pan heats up; what heats up place x on the pan is always some specific mass-energy displacement

But when we say things like "*someone's* coughing caused the pianist to hit a wrong note", "the presence of the flame caused the pan to heat up", and "the freezing of the liquid caused the vessel to shatter", what we are saying is surely not altogether wrong. There is plainly a sense in which it is true to say that *somebody's* coughing is what caused the pianist to error, a sense in which it is true to say that the presence of the flame caused the pan to heat up, and so on.

Here is how Jackson and Pettit deal with it. The presence of the flame *programs for* the heating-up of the pan. Put another way, it is *causally*

relevant without being *causally efficacious* in the heating up of the pan. What does this mean? Given that there is a flame present, it necessarily follows that something which *is* causally efficacious in the heating up of the pan will be present. The flame is constituted by the specific displacements which did the causal work. And the presence of a flame in that place guarantees that *some* such displacements will be present which will do the causal work. So even though the flame *per se* is not what does the work, the presence of the flame is causally *relevant* to the heating up of the pan: the presence of the flame guaranteed (not in a causal sense) the presence of something which would do said work.

Similarly for the other cases. What causes the glass to shatter is not the fact that the liquid inside has a certain temperature or volume. It is something more specific: the impact of certain molecules on certain other molecules (actually it is eve more specific than that). But the temperature of the glass is causally *relevant.* Given that the glass has such and such temperature, it is guaranteed (not causally, but in some other, deeper sense) that there will be the kind of thing which *is* causally efficacious in the shattering of the glass.

The fact that *somebody or other* coughed isn't what caused Horowitz to err. But that fact *does* guarantee the existence of the sort of thing which *would* do that. Given that *somebody* or other coughed, it follows that, for some specific x, *x coughed*: and the fact that x coughed *is* causally efficacious in causing Horowitz to err. So *somebody or other's coughing* is causally relevant, though not causally efficacious, to Horowitz's erring. It programs for the existence of something which is causally efficacious, even though it does not itself have that property.

Jackson and Pettit say that, if we take into account the distinction between efficacy and relevance, we can be functionalist or externalists *and* hold onto the idea that content has an important causal role. What is causally

efficacious in Smith's reaching for the phone and dialing those numbers is not the fact that Smith is thinking about Mary. But Smith's thinking about Mary is *causally relevant.* His doing so "programs for" the kind of thing that will do the trick. So content has causal relevance (though not causally efficacy).

Similarly for the functionalist-case. (For the moment, let us forget about the problems relating to differences between thinking about water/twin-water, and the like.) For the reasons we described earlier, Smith's wanting water (or being in pain or believing that there is arsenic in the glass…) cannot be *causally efficacious* in Smith's reaching for the glass: it cannot, strictly speaking, what causes Smith to do that. But it can be *causally relevant.* Smith's being thirsty "programs for" his doing that sort of thing; the presence of thirst guarantees the presence of the kinds of things which would to the causal work in Smith's reaching for the glass.

First we will deal with the functionalist case. Let us start with some platitudes. Nothing can be "causally relevant" to itself. A state of affairs cannot be "causally relevant" to its own existence. Of course, the term "causal relevance" is an invented one; so we can give it any meaning we wish. But surely it would be wrong to say describe the kind of relevance that event E has to itself as *causal.* The relation is better described as *logical* (or perhaps metaphysical). E is not *causally* related to itself; and it doesn't matter, obviously, whether we are taking about causal efficacy or causal relevance. Obviously the kind of relation Jackson and Pettit have in mind, when the talk about "causal relevance" is *some* kind of causal relation; the word "causal" is there for a reason. So nothing can be "causally relevant" to itself.

The functionalist says that x's causing E1…En is what x's having C

consists in. There is an identity here, not a causal relation. At the same time, it is clear, from a common-sense point of view, that x's having content C – e.g. x's being a desire for water – has some kind of causal hand in bringing about E1…En (in bringing it about that Smith reaches for the glass or tries to melt the icicle or whatnot). If x's having C is *identical with* x's bringing about E1…En, x's having C cannot be *causally effective* in bringing about E1…En: after all, x's bringing about E1…En cannot cause x's bringing about E1…En. By the same token, x's bringing about E1…En cannot be *causally relevant* in bringing about x's bringing about E1…En. Nothing can be *causally relevant* to itself. If the functionalist is right, then x's bringing about E1…En is *identical* with x's having C. So if the functionalist is right, then x's having C – i.e. its bringing about E1…En – cannot be causally *relevant* to its bringing about E1…En: not causally relevant, not causally effective. So the Jackson and Pettit maneuver fails.

To deal with the externalist-case, a background point is in order. Sometimes Jackson and Pettit analyze (or define) the notion of causal relevance in counter-factual terms. x is causally relevant to the occurrence of y if, in any counterfactual circumstance where x exists, there also exists something causally effective in bringing about y. The problem is that this is too generous. My striking the glass with the hammer is sufficient, in the counterfactual sense just defined, for the breaking of the glass. The same is true my of striking the glass and wearing a yellow hat and having green teeth and wearing a spacesuit…But surely my wearing a yellow hat/space-suit/having green teeth…is not in any sense "causally relevant" to the breaking of the glass. The counterfactual analysis is too broad, too expansive.

It needs to be tightened. But we don't want to tighten it too much; we don't want to say that (e.g.) the program-cause (the thing that is causally is relevant) is the *smallest* thing x such that x counterfactually guarantees the

presence of something causally efficacious. After all, the smallest thing is the mass-energy displacement itself: we don't want to shrink causal relevance down to causal efficacy.

The notion of causal relevance is, I believe, to be understood in terms of the notion of *redundancy.* Let x be the particular mass-energy displacements (constitutive of the fire) which did the trick. The flame is causally relevant because in any case where you have the flame, but where x (for some reason) didn't do the trick, the flame guarantees that some other mass-energy displacements y *will* do the trick. So causal relevance can be understood in terms of "back-up" mechanisms. In any situation where *someone or other* coughed loud enough, but where Smith (the actual cougher) did not cough, that is enough to make Horowitz hit the wrong key. Now Smith's wearing a yellow hat is *not* causally relevant, because it doesn't constitute the right kind of back-up mechanism; it doesn't amplify the range of cases where something does the trick. By contrast, the *whole* flame is causally relevant because, if the flame were smaller, that would eliminate some back-up mechanisms; it would shrink the range of mass-energy displacements that *would* step into the breach in situations where X isn't being causally efficacious. But Smith's wearing a yellow hat, the flame's coming from a green as opposed to a blue lighter – or its coming from a lighter as opposed to a newspaper – doesn't provide this sort of counterfactual reinforcement. It is inert in this respect.

The same is true of the fact that Smith's brain-state has its origin in Mary, as opposed to twin-Mary, or Mary3, or Robo-Mary, or hologram-Mary or a Cartesian demon or bent light-rays. What is causally *efficacious* is some congeries X of mass-energy displacements at various contact-points in Smith's nervous system. What is causally *relevant* are the structures that guarantee, in counter-factual situations where X didn't do the trick, some

congeries Y would. What is *not* causally relevant is anything that doesn't provide this sort of reinforcement; where that ends, so does causal relevance. Now consider the things going on outside Smith's cranium. Are any of *these* things going to step into the breach in counter-factual situations where X isn't doing the trick? Consider a counter-factual situation where X isn't doing the trick. Is the fact that Smith's brain-state originates with Mary (as opposed to twin-Mary or Mary3) going to help out? Is it going provide needed back-up? No. It is inert. It would provide back-up *only* to the extent that it led to the right kind of *brain-structure* or, in any case, the right kind of *local* phenomenon that would provide back-up.

Program-causes and Mental Efficacy

I. It seems reasonable to suppose that, in some cases, brain-states realize mental contents. (I wish the word "state" to be taken as broadly as possible – to apply to both enduring brain-structures and occurrent phenomena.) My belief that 1+1=2 is realized by certain brain-states. My belief that water is wet is realized by a different set of brainstates.

Let x be an arbitrary brain-state of mine; and suppose that x is identical with, or realizes, my belief that bread nourishes. According to functionalism, for x to realize that belief is for x to have a certain causal role. For x to have a certain causal role is for it to be caused by certain things (e.g. visual perceptions of people eating bread and subsequently appearing sated) and for x, in its turn, to cause certain other things (e.g. my choosing to eat bread when hungry). So, for functionalism, x is a belief that bread nourishes in virtue of (inter alia) the fact that x leads to my deciding to drive to the store to buy bread, to my eating said bread, to my teaching loved ones how to bake bread, and so forth.

There is an obvious problem for functionalism. If x's realizing a belief that bread nourishes *consists* in (inter alia) its leading to certain states of affairs – e.g. my choosing to eat bread, my deciding to drive to the bread-store when hungry, and so forth – then x's having that content cannot be what *causes* those states of affairs to come about. If x's having that content *is* its having a certain causal role, then x's having that content cannot *affect* or *govern* its causal role; it cannot be what brings about my choosing to eat bread or my teaching my loved ones how to bake it. In a word, if content *is* causal role, then content cannot *affect* causal role: content ends up *not* having a causal hand in contexts where, quite plainly, it *does* have such a hand. [290]

II. Another doctrine – content externalism – also seems to make content inert. Smith and Twin-Smith are atom for atom duplicates. But Smith is thinking about Mary, while twin-Smith is thinking about twin-Mary. The reason is that Smith's brain-state has one distal cause (Mary), while twin-Smith's has another (twin-Mary). The difference in distal cause makes for a difference in content.[291]

But this difference in content seems to be inert. What does all the work is *local*: the spatially and temporally local properties of Smith's brain are what cause him to pick up the phone and dial those numbers; the same is true (mutatis mutandis) of Twin-Smith. The differences in distal causes are causally irrelevant: everything that does causal work is in the here and now – is lodged inside Smith's and twin-Smith's respective crania. It is irrelevant that Smith's state was distally caused by Mary[292], as opposed to twin-Mary or hologram-Mary or Robo-Mary. Smith's state might as well have been caused by any one of them, so far as the causal properties of that state are concerned.

So to the extent that mental content is to be understood in externalist terms – to the extent that Smith and twin-Smith have qualitatively different mental contents in virtue of their differing causal liasons -- mental content doesn't do anything. Externalism thus appears to strip mental content of causal power.[293]

III. Jackson and Pettit deal with these problems in a brilliant way. First, they distinguish causal *efficacy* from what they refer to as causal *relevance.* They then attempt to show that, given this distinction, we can reconcile externalism and functionalism with the presumed causal potency of mental content.

First let us discuss the distinction between efficacy and potency. Consider

the statement:

"The presence of a flame under the container caused the liquid to boil."

Well, what really caused the container to heat up is much more specific: certain specific displacements of energy led to some other specific displacements of energy; and these displacements happened in a very specific region – a region coincident with, or vanishingly close to, the region occupied by the under-surface of the pan. Not every displacement of energy constitutive of the flame caused the increase in temperature. It wasn't really *the flame* which did the work: it was certain very specific micro-events constitutive of the flame.

Causes are much more specific than our statements typically indicate; and the things we describe as "causes" – e.g. a *flame's* being underneath the container – are not really what is doing the causal work; there is always something much more specific going on, e.g. some specific molecule (or atom or quark…) bouncing off of some other molecule (or atom…).

But when we say things like "the presence of the flame caused the pan to heat up", what we are saying is surely not altogether wrong. There is plainly a sense in which it is true to say that the presence of the flame caused the pan to heat up.

Here is how Jackson and Pettit deal with these facts. The flame is constituted by the specific displacements which did the causal work – which are causally *efficacious.* And the presence of a flame in that general region guarantees that *some* such displacements will be present which will do the causal work. So even though the flame *per se* is not what does the work, the presence of the flame is causally *relevant* to the heating up of the pan: the presence of the flame guarantees or "programs for" the presence of

something which would do said work. Sometimes Jackson and Pettit describe the thing which is causally *relevant*, but not causally *efficacious*, as the "program cause".[294] To sum up, the flame is causally relevant, but not causally efficacious. What is causally efficacious is the activity – the dance of mass-energy displacements - at the meeting point of the flame and the underside of the pan.

IV. Jackson and Pettit say that, if we take into account the distinction between causal efficacy and causal relevance, we can be functionalists or externalists *and* hold onto the idea that mental content has an important causal role. What is causally *efficacious* in Smith's reaching for the phone and dialing those numbers is not the fact that Smith is thinking about Mary. But Smith's thinking about Mary is *causally relevant.* His doing so "programs for" the kind of thing that *will* be causally efficacious in his hitting those seven numbers. So content has causal relevance (though not causal efficacy).

So as far as what is causally *efficacious* is concerned, it is irrelevant whether the distal cause of his current condition was Mary, twin-Mary, robo-Mary, or a hallucination. Thus Smith's thinking about Mary – as opposed to twin-Mary or robo-Mary – is not causally efficacious. But Smith's thinking about Mary is still *causally* relevant, as it *programs for* his dialing those seven numbers.

The same principle applies in the case of functionalism. For the reasons we described earlier, Smith's wanting water cannot be *causally efficacious* in Smith's reaching for the glass: it cannot, strictly speaking, be what causes Smith to do that. But it can be causally *relevant.* Smith's being thirsty "programs for" his doing that sort of thing; the presence of thirst guarantees the presence of the kinds of things which would to the causal work in Smith's

reaching for the glass.

IV. I believe that Jackson and Pettit have had a profound insight into causality and causal explanation. But I don't think that insight helps out functionalism or externalism in the way that they think.

Let us start with a platitude. Nothing can be "causally relevant" to itself. A state of affairs cannot be "causally relevant" to its own existence. Obviously things are relevant to themselves. But the relevance isn't *causal.*

Of course, the term "causal relevance" (as Jackson and Pettit are using it) is an invented one; so we can give it any meaning we wish. But surely it would be deeply misleading to describe the kind of relevance that event E has to itself as *causal.* The relation is better described as logical or perhaps metaphysical. E is not *causally* related to itself; and it doesn't matter whether we are talking about causal efficacy or causal relevance. Obviously the kind of relation Jackson and Pettit have in mind, when they talk about "causal relevance" is *some* kind of causal relation; therefore it is not a relation which something can have to itself -- the word "causal" is there for a reason. So nothing can be "causally relevant" to itself.

The functionalist says that x's having content C *identical with* or *realized by* x's having causal role R. So, for the functionalist, x's having C is realized by (inter alia) x's leading me to bake bread when hungry, and so forth. Since we are here dealing with a relation of identity or realization, we cannot say that the relation between x's having C and x's having causal role R is any sense *causal.* It doesn't matter whether we are talking about causal relevance or causal efficacy. Both causal relevance and causal efficacy are *causal* relations (of some kind or other); and thus do not hold between identicals or between things one of which realizes the other. So x's having content C cannot possibly be causally relevant to x's having causal role R. It is no more

possible for the one to be causal relevant with respect to the other than it is for the one to be causally efficacious with respect to the other.

Let us sum up. For the functionalist, x's having C is identical with, or realized by x's causal role R. If the relation between x's having C and x's having R is one of identity or realization, then there is a categorical ban on *any* kind of causal relation – whether one of efficacy *or* relevance – holding between the two. So the Jackson-Pettit maneuver fails to reconcile functionalism with the intuition that my believing that bread nourishes has causal potency.

V. To deal with the externalist-case, a background point is in order. Sometimes Jackson and Pettit analyze the notion of causal relevance in counter-factual terms: x is causally relevant to the occurrence of y if, in any counterfactual circumstance where x exists, there also exists something causally effective in bringing about y.

The problem is that this is too generous. My striking the glass with the hammer is sufficient, in the counterfactual sense just defined, for the breaking of the glass. The same is true of my of striking the glass *while* wearing a yellow hat and having green teeth and wearing a spacesuit…But surely my wearing a yellow hat (or having green teeth) is not in any sense "causally relevant" to the breaking of the glass.

The counterfactual analysis is too inclusive, and thus needs to be tightened. But we don't want to tighten it too much. For example, we don't want to say that the program-cause is the *smallest* thing x such that x counterfactually guarantees the presence of something causally efficacious. After all, the smallest thing that gives that guarantee is the mass-energy displacement itself. But we don't want to say that *only* that mass-energy displacement is causally relevant: surely we want to allow that the flame of which it is a part

is causally relevant; we don't want to shrink causal relevance down to causal efficacy.

The notion of causal relevance is, I believe, to be understood in terms of the notion of *redundancy* – in terms of the notion of a *back-up mechanism.* Consider the case where, as we loosely say, the presence of the *flame* caused the pan to heat up. Let x be the *specific* mass-energy displacements constitutive of the fire which, in the actual case, did the trick. The flame is causally relevant because in any case where you have the flame, but where x (for some reason) didn't do the trick, the flame guarantees that some other mass-energy displacements y *will* do the trick. So, it seems, causal relevance can be understood in terms of "back-up" mechanisms.

Suppose that the wall-paper in the kitchen where the flame occurs is yellow. That isn't causally relevant to the heating of the pan, because it doesn't amplify the range of cases where something is causally efficacious in the heating of the pan. By contrast, the flame as a whole is causally relevant to the heating because, if the flame were smaller, that would eliminate some back-up mechanisms; it would shrink the range of mass-energy displacements that *would* step into the breach in situations where x (the totality of mass-energy displacements that *in fact* did the trick) isn't being causally efficacious.

The same is true of the fact that Smith's brain-state has its origin in Mary, as opposed to twin-Mary, or Mary-7, or Robo-Mary. What is causally *efficacious* is some flurry X of mass-energy displacements at various contact-points in Smith's nervous system. What is causally *relevant* are the structures that guarantee that, in counter-factual situations where X wasn't there to do the trick, some other flurry Y would have been. What is *not* causally relevant is anything that doesn't provide this sort of reinforcement. Where that sort of reinforcement ends, so does causal relevance.

In light of this, consider the things going on outside Smith's cranium. Are any of *these* things going to step into the breach in counter-factual situations where X isn't doing the trick? Consider a counter-factual situation where X isn't doing the trick. Is the fact that Smith's brain-state originates with Mary (as opposed to twin-Mary or Mary-7) going to help out? Is it going to provide the needed back-up? No. It is inert. It would provide back-up *only* to the extent that it led to the right kind of *brain-structure* or, in any case, the right kind of *local* phenomenon that would provide back-up.

There are events which, while not themselves being causally efficacious in making Smith dial those seven numbers, do yet program for his doing so. What is it that is causally efficacious in making Smith dial (or at least choose to dial) those numbers? Answer: certain specific mass-energy displacements at, or vanishingly near, certain neural contact points. So what is causally efficacious isn't exactly some chemical (or electrical) storm in Smith's brain – some flurry of z-fibers or whatnot. The *storm* is much too encompassing an entity to be causally efficacious (just as the *flame* is too encompassing an entity to be causally efficacious in heating the pan): causal efficacy resides at (or vanishingly near) contact points; the storm *includes* things that happen at those contact points, but it also includes things that are (relatively speaking) far away from them. Nonetheless, that chemical storm – the thing that neurologists speak of as "the cause" of Smith's number-pushing muscle-spasms – in some sense *guarantees* that the right kind of events will happen at those contact points. What is the nature of that guarantee? The chemical storm *consists* of those events (and others). The occurrence of the storm guarantees the existence of the right events at the contact points because the storm *consists* of events that will do the trick. In the actual world, the storm comprises E1…En – these being the relevant events at the contact-points. But if the storm hadn't comprised specifically E1…En, it *would* have comprised

E*1…E*n, where E*1…E*n would have done the trick. It *would* have done this because the chemical storm is composed of the right kind of events – not because it would have *caused* the right kind of events. So what is *causally relevant* to, but not causally efficacious in, Smith's dialing those numbers is *not* some spatiotemporally remote state of affairs involving Mary: it is some contemporaneous state of affairs which *comprises* the causally efficacious events.

Events buried in the past (even the very recent past) – whether they involve Mary, twin-Mary, or Robo-Mary – are not, in Jackson and Pettit's technical sense of the phrase, "causally relevant" at all. They are *causes*; but they are not "causally efficacious", Jackson and Pettit's technical sense – for causal efficacy resides in what is doing the work right now. And they are not "causally relevant", in the technical sense – for they are too remote to provide the kind of counterfactual reinforcement we spoke of earlier. That sort of reinforcement is provided by brain-events which are contemporaneous with, and (relatively speaking) very close to, the events which are causally efficacious. These (relatively) local events – and not some remote state of affairs involving Mary (or Alpha Centauri) -- are the ones that will pick up the slack in counterfactual cases where the events that are *in fact* causally efficacious are truant.

References

Jackson, Frank; Pettit Philip; Smith, Michael. 2004a. *Mind, Morality, and Explanation.* Oxford: Oxford University Press.

-- 2004b. "In Defense of Folk Psychology." In Jackson, Frank; Pettit Philip;

Smith, Michael. *Mind, Morality, and Explanation.* Oxford: Oxford University Press.

-- 2004c. "Causation in the Philosophy of Mind". In Jackson, Frank; Pettit Philip; Smith, Michael. *Mind, Morality, and Explanation.* Oxford: Oxford University Press.

-- "Functionalism and Broad Content". In Jackson, Frank; Pettit Philip; Smith, Michael. *Mind, Morality, and Explanation.* Oxford: Oxford University Press.

-- 2004e. "Program Explanation: A General Perspective". In Jackson, Frank; Pettit Philip; Smith, Michael. *Mind, Morality, and Explanation.* Oxford: Oxford University Press.

-- 2004f. "Structural Explanation in Social Theory". In Jackson, Frank; Pettit Philip; Smith, Michael. *Mind, Morality, and Explanation.* Oxford: Oxford University Press.

-- 2004g. "In Defense of Explanatory Ecumenism". In Jackson, Frank; Pettit Philip; Smith, Michael. 2004e. *Mind, Morality, and Explanation.* Oxford: Oxford University Press.

Kaplan, David. 1968. "Quantifying In". Reprinted *The Philosophy of Language.* A.P. Martinich (ed.). Oxford: Oxford University Press.
--1989. 'Demonstratives'. In *Themes from Kaplan.* Edited by Joseph Almog, John Perry, and Howard Wettstein. Oxford: Oxford University Press.

-- 1989b. “Afterthoughts”. In *Themes from Kaplan.* Edited by Joseph Almog, John Perry, and Howard Wettstein. Oxford: Oxford University Press.

Putnam, Hilary. 1975. “The Meaning of ‘Meaning’”. In *Collected Papers: Volume II.* Cambridge: Harvard University Press.

Externalism and the Veil of Perception

There are two opposed conceptions about our perceptual relation to external reality. Some (e.g. Hume, Ayer) say that we perceive *representations* of external objects, and not the objects themselves: there is thus a "veil of perception" lying between ourselves and the out world. Others (e.g. Reid) say that we perceive the objects themselves.

Sometimes the former position is known as "indirect realism": there really are external objects, but we are only indirectly aware of them – we are aware of representations of them. Predictably, the latter position is sometimes referred to as "direct realism": there really are external objects, and we are directly aware of *them*, not of representational intermediaries. (In this context, I am taking for granted the existence of the external world; so it isn't worth discussing the relations of internalism or externalism to *anti*-realism.)

On the face of it, it might seem as though *externalism* is associated with direct realism, while internalism is associated with indirect realism. According to the externalism, Alpha Centauri itself is a veritable constituent of my thoughts. No Alpha Centauri, no thought. Thus, given that the thought exists, there is no need to *connect* the thought to external reality: if it exists at all, the thought exists *connected* to the external world.

By the same token, indirect realism seems to be associated with *internalism.* The internalist says that the content of our mental states is confined to what is internal to the "Cartesian theatre". So it would seem that, for the internalist, the content of our states does not, at least not directly, have anything to do with anything that lies outside the Cartesian theatre: Alpha Centauri is not *itself* a constituent of our thoughts. What is a constituent of our thoughts is, at most, something that *represents* Alpha Centauri. So internalism appears to be committed to the view that perceptual awareness is mediated by a veil of

representations. (Let us refer to this view as VOP for "veil of perceptions.)

There is some historical tendency for internalists to accept VOP – for example, Descartes accepted VOP and was an internalist, so far as he can be classified either as an internalist or as an externalist. This is not universally true. John Searle is a direct realist who is also an internalist.

I think that this debate needs to be re-examined in light of some of the points we've discussed. Thus re-examined, it turns out that, ironically, externalism is committed to a rather extreme form of VOP. A good point of entry into this debate is "Brentano's thesis":

(BT) All awareness is awareness of something.

(BT) seems reasonable enough. It is hard to imagine a "pure" awareness – an awareness that wasn't an awareness *of* this or that.

But there is a famous problem with (BT). Suppose you hallucinate a cat. What is the object of your awareness? A cat? There is no cat – at least, no *real* cat. Yet in having the hallucination, you are obviously aware. So here, it will be said, we have an awareness that is not an awareness of anything. Thus, it seems, we have a counter-example to BT.

Brentano, and others, have dealt with this by saying that, when you hallucinate a cat, you *are* aware of a cat: a non-existent (or merely "subsistent") cat. But this is obviously not good reasoning. Either there is a cat or there isn't. If there isn't, then there isn't any cat to be aware of. Either we are simply playing with labels or we are ontologizing to cover up deficits in our reasoning.

But hallucinations are no counter-examples to BT. For reasons earlier discussed, the content of your perception is existential. When you 'see Fido', as we put it, what your perception tells you is existential: *there is a creature*

with such and such properties. The perception is *made true* by some fact about Fido. But Fido is not himself a constituent the content of that perception. To think otherwise, we saw, is to confuse perceptual content with *meta*-perceptual content. Given *only* the information encoded in your perception, you have no way (no matter how intelligent you are) of knowing whether you are seeing Fido or twin-Fido or Android-Fido (or nothing at all). The only information that could help you decide that is meta-perceptual – information obtained through some source other than that very perception. What is given to you in that perception is neutral between Fido's being in front of you and twin-Fido's being in front of you and android-Fido's being in front of you. There is nothing *in that perception* itself that could possibly warrant a decision as to which, if any, of these three things (Fido, twin-Fido, Android-Fido) is the one in front of you. Thus the perception cannot deliver a message of the form…*Fido.*..(as in *Fido is looking ill* or *Fido is panting heavily*). For if the perception did deliver such a message, it would *not* leave it open, even epistemically, whether it was Fido or twin-Fido (or neither) that was in front of you. So Fido cannot be a constituent of the information conveyed to you that perception.

What is given to you in that perception is Fido-neutral. What is given to you is that there is *some* creature with such and such properties, and is thus purely existential. So *objects* are not the content of perceptions; the content of a perception is *made true*, in favorable cases, by an object; but the object is not itself lodged in that content. Objects are the objects of perception only in the attenuated sense in which "Sally met a man" is about Frank or Harry or Doug (whoever it was that Sally met). We must distinguish content from truth-maker.

If we maintain that external objects or even states of affairs *are* constituents of the information borne by your perceptions, then we end up guilty of self-contradiction. As we noted, given only the information that is *in* the perception, it is possible that you are seeing android-Fido or even that there is

no external world at all. So the content of the perception leaves it *open* whether Fido is there – whether anything is there, for that matter. That content, as we said, is Fido-neutral: *that* specifically is what it means to make the true claim, given that perception, it is epistemically possible that Fido isn't there. Given – what nobody denies -- that the perception leaves is epistemically open whether Fido even exists, we must acknowledge that the content of the perception is Fido-neutral – Fido is no part if of it. So if we say that Fido himself *is* part of the content that perception, then we are forced to say, absurdly, that Fido both is and is *not* part of the information encoded in that perception.

Uncontroversially, given only what the perception tells you, it is epistemically possible that you are seeing android-Fido, or twin-Fido. So if, with the externalist, we say that Fido is a constituent of the content of your perception, then we must say, absurdly, that the content of the perception is not part of what the perception tells you. But what is the content of a perception if not what it tells the percipient? If we separate the content of the perception from what the perception tells you – and this is exactly what the externalist must do – then we end up operating with an incoherent or vacant notion of content.. So we don't want to make that separation. But if we *don't* make that separation, then as we saw, Fido himself ends up *not* being anywhere in that content, in which case externalism is false. So to avoid incoherence, we are forced to say that the content of your perception does not include Fido at all. That content is not *Fido*, nor anything of the form…*Fido*…Rather it is, for reasons that are independently argued, some existence claim that is *made true* by the Fido.

Let us relate this point on Brentano's thesis. When you have a hallucination of a unicorn, the content of your condition is not simply *Pegasus*. It is (say) *there is a horse-like creature with a horn on its head and white fur*…The

content of your perception is not a non-existent unicorn. If there *were* a unicorn there, the content of your perception wouldn't be an existent unicorn, as we, in effect, just saw when discussing Fido. The content of your perception is given by an existence claim. What your perception is saying is: *there is a being with such and such properties.*…You don't see a non-existent unicorn. At the same time, your state is obviously *representational*: its representational content is given by an existence claim, and is not identical, even in part, with an unicorn (existing or not existing). So the existence of hallucinations doesn't threaten BT, and we don't have to ontologize to protect BT.

Once we acknowledge the fundamentally existential nature of sense-perception, it becomes clear that, ironically, externalism is a form of VOP and that internalism is not. The externalist says that the relation between a perception and its content is (epistemically) contingent. According to the externalist, your perception has Fido for at least part of its content. At the same time, everybody agrees that it is *epistemically* possible that you are seeing twin-Fido, and not Fido. In fact, it is epistemically possible that there is *no* dog (or dog-like entity) there at all. In that case, according to the externalist, the *content of your perception doesn't exist* or, at least, it has a gap in it.

The situation is dire. It is *epistemically possible,* as Descartes and Hume pointed out, that there is no external world at all. (I doubt anyone would deny this, even though few would accept the skepticism to which this fact drove Hume and Descartes.) Now the externalist identifies content with something extra-cranial. So if the externalist is right, then it is epistemically possible that your perceptions don't have *any content at all.*

Evans, MacDowell, and other externalists maintain that, where there is no object, there is no content. They are being consistent here – true to their own externalism. *If* externalism is right, then the very *content* of your perception of the dog includes that dog itself. If that is right, then the representational

content of a hallucination of a dog has a gap in it where there should be a dog. To the extent that a state is a hallucination, it is *missing* content. Given this, suppose that *all* of our waking life is a dream – this is obviously *epistemically* possible. In that case, the externalist must hold that are experiences are content-free: the (would-be) perceptual states of a brain in a vat or the victim of an Evil Demon become devoid of content.

But this is very hard to believe; in fact, it is self-contradictory. When we say that it is *epistemically* possible that a given perception is a hallucination, we mean that it is possible *given the information that is encoded in the perception.* So talk of what is epistemically possible *presupposes* that perceptions have content. Everybody, externalists included, wishes to admit that it is *an epistemic possibility that*, right now, I'm seeing twin-Fido, as opposed to Fido, or even that I'm having a hallucination and am not seeing anything at all Since everybody, externalists included, countenance the idea that certain things are *epistemically* possible relative to the perception, it follows that everybody, externalists included, are committed to the idea that the perception has a *resilient* content – a content that exists in all epistemically possible counter-factual scenarios where that perception exists: scenarios where Fido doesn't exist, scenarios where we are dupes of Evil Demons. So it makes no sense at all to say that it is epistemically possible that our perceptions have no content. When we say that such and such is epistemically possible with respect to our perceptions, we are saying precisely that the content of our perceptions doesn't rule it out. The externalist says that the content of the perception is extra-cranial. So the externalist is forced to make the false, indeed the absurd, claim that it is *not* epistemically possible that I am seeing twin-Fido (as opposed to Fido) or that I am a hallucinating.

Let us use the term "perceptual state" to describe perceptions and hallucinations alike. (So the word "perceptual" is being used as a mnemonic

device, not as a way of denoting the property of *really* being a hallucination.)

The *contentfulness* of our perceptual states cannot be doubted. If we say of some perceptual experience that it *could* be a hallucination, we are saying that *given its content*, it is an open question whether it is a hallucination or not. (This is an open question where every perceptual state is concerned.) So scepticism *presupposes* contentfulness. So if the externalist wishes to be coherent, he must *accept* the contentfulness of our perceptual states – the veridical perceptions and the hallucinations alike. The Evans-MacDowell view is that, in so far as perceptual state is a hallucination, it has a *gap* in it. A consequence of this view is that, if we are victims of Evil Demons, then our perceptual states would turn out to be altogether content-free. But we've just seen that our perceptual states *do* have content, even if it turned out they were all hallucinatory, and that it is incoherent to maintain otherwise. The epistemic possibility of hallucination presupposes the contentfulness of our perceptual experiences. Therefore the Evans-MacDowell view is incoherent, given the very notion of epistemic possibility. So to avoid incoherence, the externalist must actually hold onto the idea that the contentfulness of our perceptual experiences is not jeopardized by the epistemic possibility of global hallucination. But Ian *externalist* could not do this. For what survives sceptical doubt is precisely what is internal. So in so far as contentfulness is Demon-proof (or vat-proof) it follows that contentfulness is not external, but is inseparable from the perceptual experience itself.

For the externalist, the relation between a perceptual experience and its content becomes epistemically contingent. You can have the perceptual experience: even so, the *content* of the experience might or might not exist. After all, the *content*, if it exists, will be some external state of affairs – some star that might be long gone or might never have existed to begin with. So the perceptual experience is one thing, its content – what it *tells* the percipient – is

another. And the relation between the two is fragile. It is possible, at least epistemically, that *that* very perception might have had a very different content. So for the externalist, the perception has an external relation to its own content: the perception is a veritable intermediary, a veil; its content is on the other side of that veil. When that veil is peeled away, so to speak, the content might be there – or it might not. The root-commitment of externalism thus turns out to be the idea that perceptions are intermediaries: the content of the perception is quite separable from the perception itself. The Humean view, powerfully critiqued by Reid, recrudesces in externalism.

With the internalist the story is different. The content of the perception, for the internalist, is Demon-proof, vat-proof, and generally hallucination-proof. For the internalist, you cannot have the perceptual experience without the content. So there is no layer of perceptions *between* subject and content. Once there is a perceptual experience, there *ipso facto* is the content. In a way, there ceases to be any intervening of layer of *representations* at all. There is not, on the one hand, a perception and, on the other hand, a separable content of perception. Where the perception is, the content is as well: the perceptual experience just *is* an awareness of content. So there is not, except nominally, a distinction between the existence of the perceptual experience, on the one hand, and awareness of content. There is just one thing – an awareness of content.

Representationalism

We intuitively or pre-theoretically divide the world into two types of entities: those that are *mental* and those that are not. Emotions, beliefs, perceptions and intentions are mental. Rocks, stars, chairs, and items of clothing are not. This dichotomy plays an enormous role in everyone's ideation: in what they believe, what they feel, what they wish for, and so on. (For example, our feelings about, and conduct towards, a given thing are determined first and foremost by whether we believe that thing to have a mind -- to have perceptions, feelings, etc.) Given this, it is worthwhile to ask: What is it that all and only mental entities have in common? With regard to all those things that we categorize as 'mental', what is it that sets them apart from those things that we categorize as 'non-mental'?

One of the more promising answers is this: something is mental just in case it has the property of being non-derivatively or intrinsically[295] *representational.* Let us refer to this position as *representationalism.* So, to be clear, representationalism (as I will be using this expression) is the doctrine that the essence of the mental lies in the property of *representationality*.[296] Equivalently, representationalism is the view that all *distinctively mental properties are representational properties*[***297***]. What sets the mental apart from the non-mental is that everything in the former category is (non-derivatively) representational, while everything in the latter is *not* representational, except perhaps derivatively. (Henceforth, whenever I say 'representational' I will mean *non-derivatively* representational, except in the few cases where the context doesn't make it clear whether derivative or non-derivative representationality is being discussed.)

Representationalism was, to my knowledge, first advocated by Franz

Brentano.[298] It was held by the so-called 'phenomenologists' (Husserl, Heidegger, and Sartre, among others). It is highly popular to this day, being vigorously defended by, among others, Fred Dretske ('All mental facts are representational facts'[299]) and William Lycan ('[T]he mind has no special properties that are not exhausted by its representational properties…',[300]). Though representationalism is a popular doctrine, its merits are fiercely disputed. Ned Block, in what has become an oft-quoted statement, says of this dispute that it constitutes '[t]he greatest chasm in the philosophy of mind'.[301]

My main purpose in this essay is to prove that representationalism is *false*. For something to be mental, it is not *necessary* that it be representational (though it is probably sufficient). So although the property of being representational is *a* distinguishing mark of the mental, it is not *the* distinguishing mark.

A couple of clarificatory remarks are in order. I defined representationalism as the doctrine that all 'distinctively mental' properties are representational properties. What exactly is a 'distinctively' mental property? A *distinctly* mental property is one that *only* a mental entity could have. Any given mental entity has to occur at a particular time and, probably, in a certain place. For example, the wave of sorrow I just felt occurred at 3:00 p.m. But obviously things that are non-mental can occur at 3:00 p.m. Representationalism is not the absurd view that *all* properties had by a mental entity are representational entities. (Every mental entity will have *some* properties that non-mental entities can have: as just noted, every mental entity has spatio-temporal properties, which, of course, non-mental things can have.) Rather, representationalism is the view that any property which *only*

mental things can have are representational properties – that, to use our new jargon, all *distinctively* mental properties are representational properties.

One more clarificatory remark is in order. 'Representationalism', as I will be using this term, is not the doctrine that, given any mental entity, there is some existing object or state of affairs that it picks out. Such a doctrine would be patently untenable.[302] For presumably there is no (real) object or state of affairs picked out by e.g. a hallucination of a pink elephant. Representationalism is the view that any mental entity is representational in the sense that it is the kind of thing that (in a non-derivative way) *could* pick out an object or state of affairs. A hallucination of a dagger is representational in the sense that it *could* apply to an object. More precisely, such a hallucination is representational in the sense that it has *conditions of satisfaction*: we can speak of what the world *would* have to be like if that hallucination *were* to be veridical -- if that mental entity were, so to speak, to be 'satisfied'. (We obviously cannot speak of what the world would have to be like for e.g. a rock to be veridical: such talk – unless meant metaphorically or derivatively – would be non-sensical.) A property is 'representational' (as we will be using the term) if it is such that, by virtue of having that property, something has *conditions of satisfaction.* So representationalism is the view that all distinctly mental properties are representational in the sense that, if something has one of those properties, it *ipso facto* has *conditions of satisfaction* or – to use an alternate locution – it *ipso facto* has *representational content.* (To be clear: I will be using the terms 'representational content' and 'conditions of satisfaction' interchangeably: for something to have representational content is for it to have conditions of satisfaction. Not all authors use the term 'representational content' in quite this way. So my usage of this expression is, at least to some extent, stipulative) So representationalism can be formulated as the view that, given

any distinctly mental property – e.g. the property of being euphoric – for something to have that property is for it to have certain conditions of satisfaction/representational content.

Now if representationalism is correct – if all distinctly mental properties are representational properties – then it follows quite trivially that two mental entities cannot differ in any of their distinctly mental properties unless they differ in their representational properties. (If two mental entities x and y differ in some of their distinctly mental properties, but do not differ in their representational properties, then *ipso facto* there are some distinctly mental properties other than representational properties, and consequently representationalism is false.) This consequence of representationalism will be the focal point of much of the following discussion.

Prima facie representationalism has much to recommend it. First of all, it seems pretty clear that *only* mental things are (non-derivatively) representational. Rocks, trees, items of clothing, etc. do not represent anything, except by virtue of their association with mental processes. [303] Consequently it seems reasonable to suppose that if something is (non-derivatively) representational, then in virtue of that fact alone that thing qualifies as mental. So for something to be mental, it is probably *sufficient* that it be representational: in a word, *all representational things are mental*.

But is the converse true? Are all mental things representational? This is where the controversy begins. To be sure, *most* mental entities are indisputably representational. Perceptions, desires, intentions, beliefs, regrets, and hopes have representational content. (One cannot just *intend*; one must intend *that something be the case*. I must intend to go to the store, or to stay at home, or to write poetry. I cannot just *regret*; I must regret not having become an artist or not having majored in medieval studies.) These things all have *propositions* as their complements; and propositions have conditions of

satisfaction and thus representational content.[304]

But there are reasons to doubt whether representationalism is a correct doctrine. The most obvious problem for representationalism – though not, as we will see, the *only* problem for it – has to do with the existence of 'phenomenal content'. For a mental entity to have phenomenal content is (to use Nagel's expression) for there to be 'something it is like' to have that entity. There is something it is like to taste mustard, something it is like see a sunset; something it is like to smell strong coffee, something it is like to have a tickle or an itch. By contrast, there doesn't seem to be 'anything it is like' to believe that pi is transcendental; so this belief doesn't have phenomenal content, though its conscious manifestations presumably do. It is a matter of great controversy whether the existence of phenomenal content can be brought into line with representationalism. Why is this exactly? Remember that, for the representationalist, there is nothing to being mental besides being representational. The way it feels – 'what it is like' - to have a woozy feeling is at least a part of what it is for that feeling to be mental: the phenomenology of that experience is at least part of its being something mental – is a 'distinctly mental' property of that experience. So the representationalist is committed to something that is arguably quite counterintuitive: he must hold that the phenomenal component of a mental entity reduces in some sense to the representational component of that entity; more precisely, he must hold that what it is like to have M is a function entirely of M's representational content. So suppose that M is your olfactory perception of strong coffee: the representationalist must hold that 'what it is like' to experience M is a function entirely and exclusively of M's having a certain representational content – in effect, of M's being rendered veridical by the existence of certain states of affairs, and rendered non-veridical by the non-existence of such.

So the representationalist must hold the following: the *phenomenal* content of a mental entity is determined wholly by its *representational* content. Now *if* this is true, then two entities cannot *differ in phenomenal content without differing in representational content:* to use the current jargon, phenomenal content 'supervenes' on representational content.

To sum up, the representationalist must hold that the phenomenal supervenes on the representational -- i.e. that if two mental entities differ in terms of what it is like to have them (in terms of their 'phenomenal content') then they must *necessarily* differ in terms of what their conditions of satisfaction are (in terms of their 'representational content'). [305]

Before we examine the merits of representationalism, we must distinguish it from a very closely related doctrine – a doctrine we will refer to as *intentionalism*. Intentionalism is the doctrine the phenomenal character of an experience[306] does indeed supervene on the representational content of that experience.[307] In other words, intentionalism is the position that the phenomenal content of an experience is *wholly determined* by its representational content , and thus that phenomenal content (at least after a fashion) reduces to representational content. It may seem that intentionalism coincides with representationalism. But, as we will see in a moment, this is not the case: intentionalism is a narrower doctrine than representationalism.

Intentionalism might seem to have bearing *only* on phenomenally pregnant mental entities (on what we are calling 'experiences': see endnote 12). But, by implication, it concerns *all* mental entities, not just phenomenally pregnant ones. More specifically, if intentionalism is right, it follows that given *any* two mental entities, those two entities cannot *differ* in their phenomenal content unless they differ in the representational content. (So intentionalism says, by implication, that e.g. a non-occurrent memory

that has representational content R must have the same phenomenal content as an occurrent perception that has representational content R.) Why exactly does this follow? If the phenomenal content of an experience M1 is *wholly* a function of its representational content, that means, among other things, that M1's having the representational content that it has is *sufficient* for its having the phenomenal content that it has. And from this it obviously follows that, given any other mental entity M2 that has the same representational content as M1, M2 has satisfied a condition that is sufficient for its having the same phenomenal content as M1; so that M2 must therefore have the same phenomenal content as M1.

Perhaps the following will clarify the point just made. Suppose that M1 is an experience that has representational content R and phenomenal content P. Intentionalism, by definition, says: 'whatever phenomenal content M1 has, it has that phenomenal content *entirely in virtue of its having the representational content that it has.*' So intentionalism says that the *one thing* about M1 in virtue of which it has P is that it has R: having R is *sufficient* for having P. (If having R were not sufficient for having P, then clearly M1's having P wouldn't be a function *wholly* of its having R.) So, by implication, intentionalism says that *any* mental entity that has R also has P. So intentionalism is committed to the view that, if (e.g.) some non-occurrent memory has R, it also has P. To sum up, intentionalism requires that given *any* two mental entities, those entities cannot differ in phenomenal content unless they differ in representational content; and therefore that if they have the same representational content, they also have the same phenomenal content.

How is intentionalism different from representationalism? Intentionalism says only that the *phenomenal* properties of a mental entity are wholly determined by its representational properties. Representationalism, on the

other hand, says that *all* of the distinctly mental properties of a mental entity are determined by its representational properties. It follows (as we've noted) that representationalism is committed to the view that two mental entities cannot differ in *any* of their distinctly mental properties (not just their phenomenal properties) without differing in their representational content: it thus holds that *all* distinctly mental properties supervene on representational properties. By contrast, intentionalism is committed only to the view that two mental entities cannot differ in their *phenomenal* properties without differing in their representational properties.

That said, it might appear that intentionalism and representationalism are, in effect, the same doctrine. For it might seem that the only distinctly mental properties that *might* be non-representational are phenomenal properties. So it might seem that, if the phenomenal properties of a mental entity content could be shown to supervene on (and thus, in some sense, reduce to) is representational properties, then representationalism would *ipso facto* be proven correct. (In other words, It might seem that if phenomenal content could be shown to supervene on representational content, then representationalism would *ipso facto* be proven correct.) We will see (in section IV) that, in fact, this is *not* the case : even if *intentionalism were* correct – even if the phenomenal supervened on the representational -- it would *still* not follow that *representationalism* was correct: for representationalism to be correct it is necessary *but not sufficient* that intentionalism be correct. On the other hand – what is a consequence of the last point – if intentionalism is false, then so *a fortiori* is representationalism.

My primary purpose in this paper is to show that intentionalism and (therefore) representationalism are false. Given this, the plan of the present paper is as follows. In sections II and III, I will attempt to show that intentionalism is false. (Of course, the arguments set forth in these sections, if

cogent, will indirectly show that representationalism is false.) In section IV I will set forth an argument that (if cogent) *directly* shows that representationalism is false. (In other words, the argument in that section will not show that representationalism is false *by way of showing that intentionalism is false*; the argument there will bear directly on representationalism.) In section V, I will try to identify the mis-step in an argument for intentionalism given by Professor Alex Byrne.[308]

One caveat is in order. There may be various ways of weakening intentionalism, so as to make it immune to the kinds of arguments I will set forth. My arguments will bear – not against every weakening of intentionalism – but against intentionalism *tout court*; that is, against *un*restricted intentionalism: the doctrine that, for *any* two experiences e and e' (regardless of e.g. what, if any, perceptual modalities they respectively belong to) they be they cannot differ in phenomenal content unless they differ in representational content. It seems to me that *un*restricted intentionalism is the most (perhaps the only) significant kind of intentionalism: for to admit that only a restricted kind of intentionalism holds is to admit that there *can* be phenomenal diversity where there is representational identity. And to admit this is to admit that there is *something* to having intentional content *other* than having representational content: and this, of course, scuttles any hope of a strictly representational theory of mind (or even of consciousness). Notwithstanding this disclaimer, I will set forth an outline of an argument to the effect that some of the more plausible ways of weakening intentionalism – so as to immunize it from the kinds of assaults waged in his paper – are not viable. (I will do this in section V; and my discussion of this problem will be connected with my evaluation of Byrne's argument.)

Although not all of this paper will be concerned with intentionalism, Byrne's just mentioned argument for that doctrine crystallizes and exposes

themes that will be present at every phase of our inquiry. So we would do well, I think, to state Byrne's argument right now. That argument takes the form of two premises and a conclusion (this is how he himself presents it). I quote:

Premise A

Assume the following: A subject enjoys two consecutive experiences: e, that ends at t, and then e*. e and e* differ in phenomenal character. The subject is *competent*: if she tries to retrieve information (or *mis*information) in memory, then her attempt will be successful, and she has no other cognitive shortcoming. The subject tries to detect a change in phenomenal character after *t.* The subject's memory is perfect: all the (mis)information supplied to her by *e* is available when she later enjoys e*.

Then, solely on the basis of her current experience e* and the memory produced by her past experience e, the subject will notice a change after *t*: if she gives an accurate report when e* is over, she will say that there was a change in phenomenal character after t. (pp.207-208)

Premise B

Assume that a subject enjoys an experience *e* that ends at *t* and then experience *e**, and that after *t* the subject notices a change in phenomenal character, solely on the basis of her current experience *e** and the (perfect) memory produced by her past experience *e.* Then the way things seem to the subject when she enjoys *e* differs from the way things seem when she enjoys *e**. That is the [representational] content of *e* differs from the content of *e**. (p. 210)[309]

Conclusion C

If a competent subject whose memory is perfect and who tries to detect a change in phenomenal character enjoys consecutive experiences *e* and *e** that differ in phenomenal character, then the content of *e* differs from the content of *e**. (p. 213)

Premise A is innocuous. All it really says is: If someone has two experiences e and e* that differ in phenomenal character, then that person will remember them as differing in phenomenal character, if he is 'competent', i.e. if his memory is accurate.

Premise B is *not* innocuous. Indeed, it contains the essence of Byrne's argument. (Arguably, it *is* Byrne's argument.) B basically says this. Suppose that some subject S has two experiences e and e*, and that e and e* differ in phenomenal content. In that case, the way the world seems to S in virtue of his having e differs from the way the world seems to S in virtue of his having e*. Of course, this is supposed to hold for *any* two experiences e and e* that differ in phenomenal content. So there can be no difference in phenomenal content without a concomitant difference in how the world seems to the subject – without a difference in *representational content.*

If I understand it correctly, conclusion C says, essentially, the following. Again, let S be some subject who has experiences e and e*, and suppose that e and e* differ in phenomenal content. Given premise B, e and e* differ in representational content. Therefore, if S is competent, he will *remember* e and e* as being different in representational content. Of course, C does follow from premises A and B.

The really controversial move in Byrne's argument is premise B. (Again,

A is non-controversial; and C non-controversially follows from A and B. So we needn't focus too much on A or C.) B makes a strong claim: where two experiences are concerned, there can be no difference in phenomenal content without a difference in representational content. Now if Byrne provided no argument for B – if he just took its truth for granted – then his argument would almost be question begging. But Byrne does provide an argument for B. It is as follows.

Suppose you are looking at a blue chip and, what is more, that you are perceiving it *as* blue.[310] Your perception of the blue of the chip obviously involves your having a certain kind of phenomenal content. (The perception of color clearly involves the experiencing of phenomenal content: there is something it is like to see an object as being red or blue.) Now imagine that dimension of your perception – viz. the dimension of your perception that mediates your awareness of the blue of the chip – as changing; i.e. imagine that that bit of phenomenal content is replaced by some other, qualitatively different bit of phenomenal content. In experiencing such a change in the phenomenal character of your experience, you are *ipso facto* experiencing a change in *how* the chip is given to you; more specifically, you are *ipso facto* experiencing a change in what color the chip is given to you as having. So with regard to your perception of the color of the chip, it is not possible for there to be a change in the phenomenology of that experience without there being a concomitant change in what that experience represents –- i.e. without there being a change in respect of what color the chip is represented to you as being. In general, a change in the phenomenal content of an experience necessarily involves a change in the representational content of that experience. So the phenomenal content of an experience is wholly determined by its representational content.

A non-visual example may make this point more starkly. Suppose that you

are tasting a drop of honey. Obviously, your having this experience involves your experiencing a certain kind of phenomenal content. Now suppose that that bit of phenomenal content changes – i.e. is replaced by a qualitatively different bit of phenomenal content. *Ipso facto* the way that the honey is represented to you has changed; more specifically, the taste that the honey is represented to you as having has changed. It is not possible for the phenomenal dimension of your gustatory experience of the honey to change without the representational dimension changing as well: if the former is changed, then so *ipso facto* is the way the honey is represented to you: in virtue of the said phenomenal change, there is a change in the *taste* that the honey is represented to you as having. So there can be no change in the phenomenal dimension of an experience without a change in its representational dimension: the former supervenes on the latter. Hence the phenomenal content of an experience is *wholly* determined by the representational content of that experience.

II. This section will provide just one reason to think that a difference in phenomenal content doesn't necessarily correspond to a difference in representational content, and consequently that both intentionalism and representationalism are false.

Suppose that you are having a veridical perception of a certain car that is a certain shade of red. Let P be this perception. Presumably, the redness of the car is represented in P by some bit of phenomenal content (by some kind of visual sensation). Surely you can have *beliefs* about the colour of the car. In fact, you can have *non-occurrent* beliefs about the colour of the car. Let B be some non-occurrent of belief of yours about the colour of the aforementioned car. Non-occurrent beliefs (by definition) don't have any phenomenal content. So B doesn't have any phenomenal content. At the same

time, there is something in B – some feature or constituent of B[311] - which represents the colour of the car. That something cannot itself be a bit of phenomenal content. (Being a non-occurrent belief, B doesn't have phenomenal content.) So it follows trivially that something *other* than phenomenal content represents the colour of the car, at least where B is concerned.

This establishes that two mental contents can have the same representational content without having the same phenomenal content. There is some constituent of P which represents the redness of the car, and that constituent has phenomenal content (in fact, it *is* a bit of phenomenal content). At the same time, there is some constituent of B which represents that same shade red, and that constituent has *no* phenomenal content. So these two bits of mental content have exactly the same representational content – they both represent the specific shade of red– but they differ in phenomenal content. (One of these bits of mental content has phenomenal content, while the other does not; so they clearly *differ* in phenomenal content.) And this, as we've seen, is inconsistent with intentionalism. And, because phenomenal properties are distinctly mental properties, it is inconsistent with representationalism.

The argument just given can be generalized. It seems that 'standing' or 'non-occurrent' mental entities (e.g. my belief that pi is irrational) do not have phenomenal content. (They have manifestations in consciousness that have phenomenal content. But they themselves do not.) *A fortiori* non-occurrent memories have no phenomenal content. You have countless memories stored away in the depths of your psyche. Surely they do not all have phenomenal content: there is not something it is like to have each and every one of them. (In fact, in so far as they are non-occurrent, *none* of them has phenomenal content.) Sense-perceptions, of course, *do* have phenomenal

content (at least conscious sense-perceptions do). As we have noted, nothing that has phenomenal content can have the *same* phenomenal content as something that has no phenomenal content: this is vacuously, or trivially, true. So, in particular, no conscious sense-perception can have the same phenomenal content as any non-occurrent memory. Consequently, the intentionalist must hold that no non-occurrent memory can have the same representational content – can encode the same information – as any (conscious) perception. (For, as we've noted, the intentionalist cannot allow that mental entities differ in their phenomenal content without differing in their representational content.)This means that, if intentionalism is correct, any sense-perception encodes some information that simply *cannot be remembered*; that, as a matter of principle, cannot be stored in a non-occurrent memory. But this is highly counter-intuitive.

Permit me to clarify this argument. It seems pretty clear that, because of our cognitive limitations, we cannot remember *everything* that we experience: few, I think, would deny this. But we also tend to think that, given any particular conscious experience, the content of that experience could, at least in principle, be stored in memory. As we've just seen, if intentionalism is correct, then – for reasons of *principle*, not reasons having to do with our cognitive limitations -- given any conscious experience, the content of that experience can*not* (in its entirety) be stored in memory. So intentionalism is really committed to the idea that some information, by its very nature, cannot be *stored*; that some information can be housed *only* in consciousness and, thus, that it can be housed only in *one* medium. This idea would seem to be quite counter-intuitive. It would seem that any bit of representational content can be stored in different media; can be transferred from (e.g.) the medium of sense-perception to the medium of non-occurrent memory. To be sure, as we noted a moment ago, a person's memory banks are limited and they probably

cannot store *everything* that that person has sense-perceived. But, if the argument just given is cogent, then intentionalism is committed to the idea that *as a matter of principle* nothing with a phenomenology – in particular, no conscious perception – can have the same representational content as something lacking a phenomenology – in particular, a non-occurrent memory. So intentionalism is committed to the view the content of your current visual perception cannot, at least not in its entirety, be stored in memory: some of the information borne by that perception simply *cannot* be remembered. And (if intentionalism is correct) the same is true of *every* conscious experience. But this idea is radically implausible and possibly even incoherent.

An intentionalist might respond to this argument with the following[312]:

Let C be some conscious sense-perception. And let N be a non-occurrent memory of C; and suppose that N is as good a memory of C as is metaphysically possible. As you've pointed out, N and C don't have the same phenomenology. And it follows – as you've also made clear – that *if* N and C had exactly the same representational content, then intentionalism would be false, given that intentionalism cannot allow a difference in phenomenology that doesn't correspond to a difference representational content. But maybe the truth is that N and C *don't* have quite the same representational content: maybe, in general, no conscious sense-perception has *quite* the same representational content as any non-occurrent memory. To be sure, N and C *overlap* in representational content; but, perhaps, they don't quite *coincide* in representational content.

If this is the case, then there is a way out for the intentionalist. For maybe this slight divergence in representational content accounts for the massive divergence in phenomenal content. To be sure, the divergence in phenomenal

content seems out of proportion with the difference in representational content. (The latter is relatively small; the former is large – indeed, it might even be said to be total.) But there is no *a priori* reason why the relationship between the two should be linear. In fact, there is independent reason to believe the kind of relationship in question is *not* linear. A thing's aesthetic properties supervene on its natural properties. So two performances of a certain work of music cannot differ in their aesthetic properties unless they differ in their natural (basically, acoustical) properties. But a slight difference in natural properties might correspond to a massive difference in aesthetic properties. Suppose that P1 and P2 are both performances of the Moonlight Sonata. P1 and P2 are very similar, in their subvenient (natural) properties, except that the piano involved in P1 is ever so slightly out of tune and the performer mis-hit one or two key notes, while the piano involved in P2 is in perfect tune, and the performer didn't mis-hit any notes. P1 may be absolutely impossible to listen to (especially to someone with perfect pitch), while P2 will be divine. So here we have a case of a slight difference in subvenient properties corresponding to an almost total difference in supervenient properties. Another analogy might clarify this. It is a clear that thing's biological properties (if it has any) supervene on its physical properties; so that creatures cannot differ in biological properties without differing in physical properties. Given this, suppose O1 and O2 are two organisms that are physically exactly a like, except that one of them has cancerous cells in some part of the brain responsible for a lot of vital functions. The strictly *physical* difference between O1 and O2 will be miniscule. But the biological difference is nearly total: one of the organisms is flourishing, while the other is on the brink of non-existence. Again, a minute difference in subvenient properties can correspond to an almost total different in supervenient properties. So the slight representational difference

between N and C is more than enough to guarantee a massive phenomenal difference, without jeopardizing the claims of intentionalism.

This response is a total failure. To fix our ideas, let us suppose that C is a visual perception of a bright red apple, and that N is a non-occurrent memory of that perception. (Also suppose that N is a relatively accurate and exhaustive memory.) As we've noted, the intentionalist must hold that the total divergence in phenomenal character between N and C is due to some marginal difference in representational content. But if we focus on the mass of representational content that N and C unquestionably *do* have in common – if we concede that they differ, however marginally, in representational content, and we confine our attention to the area of representational overlap between them – the data we are confronted with runs counter to intentionalism. Consider the bit of phenomenal content that, in C, is associated with its being a perception of something bright red; and let P be this bit of phenomenal content. The intentionalist must say that P's presence in C is due *wholly* to C's being a perception of something bright red. Now obviously N contains that same bit of representational content: so why doesn't N have the corresponding phenomenal property?

The force of this argument may be more apparent if it is stated slightly differently. If we confine ourselves to the representational content that N and C *do* have in common – and everyone must agree that there is a mass of it -- it is clear that the intentionalist must hold that, *where C is concerned,* this representational content corresponds to a mass, a deluge, of phenomenal content. So how is it that, where N is concerned, this same representational content is so utterly neutralized as regards its ability to subserve the same phenomenal content? Even if we concede that C has some extra little bit of representational content that N lacks, how could this fringe of

representational content simply cancel out the power of the remaining bit of representational content to subserve phenomenal content?

Another possible intentionalist counter-response to my argument is this:

A non-occurent memory is really just a disposition to have certain occurrent experiences, that is, to be in certain kinds of phenomenally pregnant states. Now if indeed non-occurrent memories are dispositions to have certain phenomenally pregnant experiences, then (in effect) non-occurrent memories *do* have phenomenal content.[313]

I concede that, if non-occurrent memories are just dispositions to have phenomenally pregnant experiences, then (in effect) non-occurrent memories do (at least after a fashion) have phenomenal content. (Of course, technically dispositions *per se* do not have phenomenal content. But this is a subtlety which, I concede, probably doesn't *ultimately* have much bearing on the objector's point.) I concede that *if* this analysis of what non-occurrent memories were correct, it would undermine that argument of mine for intentionalism. And I also concede that I cannot decisively refute that analysis within the confines of this particular inquiry.

What I can say is this: the objector's point – while, perhaps, a reasonable enough dialectical manoeuvre – commits him to some extremely dubious psychology and presupposes the falsity of some extremely hardy and fecund research programs in psychology. If cognitive science and psychoanalysis are to be given any credence (some say they aren't: but *if* they are), then there is much more to a non-occurrent memory (i.e. a memory which is not a conscious reminiscence) than a disposition to have certain conscious experiences. There is certainly some reason to believe that, at a level far

below that of consciousness, mental contents – including, non-occurrent memories – interact and generate new mental contents; and that their representational contents are very much implicated in these transactions. A banal example: before you go to sleep you non-occurently remember P and you non-occurently remember that if P then Q; but, for whatever reason, you haven't formed the belief that Q. You wake up believing that Q. Presumably the two aforementioned beliefs were 'inferentially' integrated and, obviously, this integration was sub-conscious; and, even more obviously, the representational contents of your non-occurrent memories were implicated in this integration. In any case, this is a plausible enough explanation. (In my view, it is inconceivable how anyone could deny that this sort of sub-conscious ideation occurs. But I acknowledge that the idea there is sub-conscious ideation has been vicious attacked and that, consequently, what I'm saying has as conditional character.) Now the representational contents of your belief that P and your belief that if P then Q may heavily overlap with the representational contents of sense-perceptions. Your belief that P may be your belief that green apples have *that* appearance (where the appearance in question is visual) and your belief that if P then Q may be to the effect that if an apple has *that* appearance one day, then it will have *that* other appearance ten days later, if it is not refrigerated. (The two 'that's here are, of course, supposed to correspond to different visual demonstrations and thus to different sense-perceptions.) Nonetheless, even though your belief that P and your belief that if P then Q have a great deal of representational overlap with perceptions, nonetheless they themselves may be non-occurrent (indeed, one has countless non-occurrent to the effect that such and such looks like *that* where the 'that' corresponds to a visual, olfactory etc. presentation); and - - this is the rub – the contents of such non-occurrent memories can, at least by the lights of some plausible psychological models, interact at a sub-conscious

level. So there is certainly excellent reason to believe the objector's analysis of non-occurrent memories (viz. that they are just dispositions to have certain conscious experiences) does not do justice to the psycho-dynamics of such things.

III. This section will set forth another argument that is directed specifically against intentionalism. Obviously perceptions that are associated with different sensory modalities have different phenomenal contents. (Phenomenally, visual perceptions are different from auditory perceptions; tactile perceptions are different from auditory perception and visual perceptions; and so on.) Now some bits of information can be given to one through different sensory modalities. I can learn that a certain object is spherical through sight *or* through touch (and a bat could learn this through hearing – through sonar – and a mole could learn this through smell). But surely no visual perception of mine has *any* phenomenal overlap with any tactile perception of mine. But, it would seem, if intentionalism is true, then there ought to be some overlap, given that there is overlap (if not identity) in terms of representational content.

The intentionalist could respond to this as follows:

Suppose that T is a tactile perception of the ball's roundness and that V is a visual perception of the ball's roundness. T and V (by hypothesis) *overlap* in representational content. But, inevitably, they will not *coincide* in representational content. For example, T will give the subject information about the firmness and the temperature of the ball, whereas V will not provide such information; and V will provide the subject with information about the ball's color, whereas T will not provide *this* information. So we have a divergence in representational content; and maybe this divergence

accounts for the divergence in phenomenal content.

We have, in effect, already seen how to respond this sort of criticism. By hypothesis, T and V overlap, to a large degree, in representational content. But T and V plainly have *no* phenomenal overlap at all. No visual perception is phenomenally at all like any tactile perception. But given how much T and V overlap in representational content, we'd expect there to be at least minimal phenomenal overlap between them. This argument can be made more explicit. Let R be the bit of representational content that T and V unquestionably *do* have in common. (So R is information to the effect that the ball is round.) Now where *T* is concerned, R corresponds to one kind of phenomenal content; whereas where *V* is concerned, R corresponds to a very different kind of phenomenal content. The kind of phenomenal content you have in virtue of *seeing* the roundness of the ball is very different from the kind of phenomenal content you have in virtue of *feeling* the roundness of the ball. So here we have a certain bit of representational content – viz. R – that is subserving quite different phenomenal contents. And this, of course, is just what intentionalism says cannot happen.

I should consider another objection to this argument:

These days a distinction is widely made between ''propositional' and 'non-propositional' content. [314] The idea is roughly this. Consider an arbitrary proposition e.g. *Bob is tall and Fred is short.* This is obviously neatly decomposable into discrete parts (*Bob, __is tall, Fred* etc.) Put in positive terms, this proposition is obviously built out of 'minimal' representational units; there is a unique and correct decomposition of this proposition. Now consider a visual perception (or a photograph). This is,

quite plainly, *not* divisible into minimal representational units; and, what is connected, the representational content of such a thing – i.e. what it represents – is not decomposable (in any natural way) into minimal units: is not decomposable into units like *Bob* and *__is tall* etc. (Propositions are 'digital'; visual perceptions are analogue.) So the representational content of a visual perception cannot be embodied in a proposition: the representational content of a visual perception is, at least in part, non-propositional (such that it cannot be embodied in a proposition).

Now you've said a great deal about *propositional* representational content. But you've said nothing about non-propositional representational content. Your argument was essentially: one's visual perception of the ball has a great deal of representational overlap with one's tactile perception of the ball – specifically, they both bear the representational content *the ball is round* – but there is no phenomenal overlap corresponding to this representational overlap. Now the representational overlap is *propositional*; both visual perceptions bear the proposition *the ball is round.* So you've made some case that, where propositional representational content is concerned, intentionalism is false, i.e. that there may be phenomenal diversity where there is *propositional* identity. But you've left open the possibility that intentionalism holds where *non*-propositional representational content is concerned; i.e. you've left open the possibility that, if two mental entities coincide in non-propositional representational content, then they coincide in phenomenal content.

There are a couple of things to say in response to this: some obvious, some less obvious. Let's start with the obvious – with what the objector himself concedes. Propositional content is *a* kind of representational content.

Given what we've already said, it seems pretty clear that perceptions with non-overlapping phenomenologies can overlap in *propositoial* content and thus in *some* kind of represnetaitonal content. (To use another example: an olfactory perception can tell you *the fish is rotten* and a gustatory perception tell you the same thing, notwithstanding that there is no phenomenal overlap between the two perceptions.) So this by itself is enough to refute intentionalism: *any* identity in representational content —whether this content is propositional or not – that subtends a diversity in phenomenal content is enough to refute intentionalism. This obvious point constitutes the bulk of my counter-response.

Now *given only what I've said* it is possible that intentionalism is true of non-propositional content; I do concede that the argument I gave above is subject to this limitation of scope. And I cannot think of any neat argument that would show that intentionalism doesn't hold for non-propositoial content. But I can make a couple of points that, while proving nothing, have *some* force against even this restricted kind of intentionalism. (What I'm about to say consists, I fully admit, of highly contentious speculations. I am permitting myself these contentious speculations since – having shown that intentionalism is false where *a* kind of representational content is concerned, viz. propositional content – I have made some case that *un*restricted intentionalism is untenable.)

First of all, with regard to those who *do* distinguish between propositional and non-propositional content, it is pretty clear that they are on to *some* kind of important distinction. But the distinction is, I submit, a distinction between *ways* of representing things; it is not a distinction in *what* is represented. Put another way, it is a distinction between ways of 'packaging' information: what is in the packages doesn't vary. There is only one kind of represnetaitonal content; but there are different ways of

transmitting this one kind of content.

Once again, consider an arbitrary visual perception. (Supposing that, indeed, some representations have non-propositional content, visual perceptions are paradigm cases of such representations; so we might as well focus on them.) Uncontroversially, the represnetaitonal content of a visual perception can *to some degree* be put into words and, thus, embodied in a proposition. People can tell you what they saw, at least up to a point ('I saw a man in an orange shirt smoking a pipe…') So obviously the represnetaitonal content of a visual perception overlaps *to some extent* with the representational content of a sentence ('I saw a man in orange shirt…'): a paradigmatically 'digital' representation. Again, I do not think this is controversial: who could possibly maintain that the representational content of a visual perception is utterly ineffable – is totally incapable of being embodied in a sentence? Surely when I say 'I saw a red apple sitting on a desk…' I am succeeding in transmitting *some* of the representational content of a visual perception of mine. (Now a hardliner might say that the representational content of e.g. my perception of the man in the orange shirt cannot *to any degree* be embodied in a proposition; that when I give a verbal report of what perception tells me, I am really only giving you some kind of analogue of some bit of information that is encoded in that perception, but that I am not really giving you any of that information *in propria persona.* Again, the obvious problem with this line of thought is that, if it is taken seriously, it forces us to take the obviously false position that, strictly speaking, our perceptions are completely and totally incommunicable. But that isn't so. In any case, I will proceed on the understandable assumption that it isn't so.)

Now the next step is also, I think, fairly uncontroversial. Visual representations are, in an important sense, homogeneous.[315] When we

consider a visual perception, there is no reason to believe that it has two, totally different kinds of representational content; visual perceptions seem to be (in this respect) 'seamless'; the representational content of a visual perception seems not to be patchwork of totally different kinds of content – of propositional and non-propositional content. Visual perceptions are not cobbled together out of discrete parts. Now if analogue representations had *both* propositional content and some other totally different kind of content, then, presumably, there would be a very big 'seam' in them; they would then have a 'cobbled together' quality which they obviously don't have. To clarify this: what it is for something to be an analogue representation is for its represnetaitonal content to have a kind of seamlessness, a kind of imperviousness to neat separation. Now obviously if a visual perception had two totally different kinds of content – propositional and other – then it would be about as capable as anything of just this kind of decomposition. So, to sum up, there is certainly a lot of intuitive support for the idea that the analogue representations (e.g. visual perceptions) have but *one* kind of representational content.

So the facts we have to work with are as follows. (i) The content of an analogue (e.g. a visual) representation can, at least up to a point, be expressed by a sentence and thus, presumably, embodied in a proposition. (ii) An analogue representation does not bear two, completely distinct kinds of represnetaitonal content.

Now given both (i) and (ii), it seems to follow that the represnetaitonal content of a visual perception is *in its entirety* propositional. But this conclusion is in deep tension with the following fact, viz.: (iii) There is, plainly, *some* kind of important distinction between analogue (e.g. visual) and digital representations (e.g. propositions): the latter are, whereas the former are not, decomposable into minimal represnetaitonal units.

The best way to make sense of these three facts is, I would suggest, as follows: the difference between propositions and e.g. visual representations is not a difference in *what* is represented but in *how* it is represented. Each of (i)-(iii) is correct. The difference between e.g. a sentence and a visual perception lies not in what is represented, but in how what is represented is *packaged.* A sentence tidies everything up, putting it in discrete parcels. A visual representation leaves it, as it were, strewn all over the floor. A proposition is not a special kind of information; it is a special *packaging* of information.

I think that this line of thought does less violence to our intuitions than the alternatives. First of all, we *do* want to say that the information borne by a visual perception can (at least up to a point) be embodied in a proposition. But we *don't* want to say that the representational content of a visual perception is a proposition (or conjunction of propositions), for the simple reason that propositions are, whereas the contents of visual representations are not, neatly decomposable into minimal units. (Surely the proposition *Bob is smarter than Fred* is decomposable into discrete parts while the representational content of a visual perception of a forest is not thus decomposable.) We also probably don't want to say that a visual perception has two wholly distinct kinds of representational content: propositional and other. Given these data, the only way to go is to say: A visual (more generally, an analogue) representation bears only kind of representational content; and this content *can* be packaged in a proposition, but it doesn't have to be.

So it is true that sentences bear propositions while visual perceptions do not; but it is also true that the content of a visual perception can be embodied in propositions. If propositions were *one* kind of represnetaitonal content while the representational contents of visual perceptions were some totally

different kind of content, this would not be possible. But propositions are not a kind of content; they are a way of packaging content. Visual representations don't package content in this way; but what they do package can be packaged propositionally.

Now if this is even remotely on the mark, then there are two quite distinct dimensions to the 'representational load' of a representation: there is the representational content of that representation; and there is also the *packaging* of that content. Now the packaging of this content has too often been identified with the representation itself. But this is a mistake. A visual perception of a round ball and the sentence 'that ball is round' obviously overlap in what they represent; but the way they package this common content differs. So what is borne by a sentence is not *just* representational content, but a certain way of packaging this content; and the same is true of a visual perception. (The sentence itself is not identical with the packaging. The sentence 'Bob is intelligent' packages the same representational content in the same way as the sentence 'Bob is smart'. But these sentences are obviously different.) But what it is that is packaged is not a proposition: and that is why visual representations – and, more generally, analogue representations: representations which do not have propositions for the representations contents – can have their contents embodied in propositions.

Now let us bring these thoughts to bear on the topic of intentionalism. We have seen, I think, that where *a* kind of representational content is concerned, two perceptions can share that content and yet have *no* phenomenal overlap. Now if – as we have just seen *some* reason (however minimal) to believe – there is only *one* kind of representational content, then – what follows trivially - if intentionalism is false for *a* kind of representational content, it is false for *all* of them.[316]

IV. In this section I will not be directly concerned with phenomenal content. I will show that mental entities can differ in some of their distinctly mental properties -- though not necessarily their *phenomenal* properties -- without differing in their representational properties (i.e. in their representational content). So my target here is representationalism, not (except perhaps indirectly) intentionalism.

Prima facie it seems that different *propositional attitudes* can have the same representational content. I can doubt whether Mary is coming home, believe that Mary is coming home, suppose that Mary is coming home, hope that Mary is coming home, fear that Mary is coming home, and so on. Basically, I can have various different *attitudes* towards the proposition *Mary is coming home.* In general, given any proposition P, one can have various different attitudes towards that proposition. This fact is of singular importance in connection with the question whether representationalism is correct. The intentionalist holds that there is nothing to the having of any mental content *besides* the grasping of representational content. So, in particular, the representationalist must hold that, given any propositional attitude, there is nothing to one's having that propositional attitude above and beyond one's grasping of some representational content. Basically, the representationalist must assimilate the *attitudinal* component of propositional attitudes to their representational or propositional component.

But it is not hard to show that, for any propositional attitude A (with the possible exception of merely *contemplating* some proposition), there is no thing x (no proposition, no object: generally, no bit of representational content) such that, in order for that person to be in A, it is enough that x be represented to that person (i.e. that that person grasp x). Basically, in order for someone to have a certain propositional attitude, it is not in general

enough that something be represented to him: not enough that he grasp some representational content -- the *attitudinal* component of a propositional attitude cannot, at least not in general, be reduced to its *representational* component.

The argument for this is fairly simple. Suppose that I fear that Mary will come home. Is there any proposition P such that, *merely* by virtue of P's being represented to me – by virtue of my grasping P – I *ipso facto* fear that Mary will come home? No. What proposition would this be? I can certainly grasp the proposition *Mary will come home* without fearing that Mary will come home. Even if there is nothing that would cause me more horror than to believe that Mary is coming home, my merely *grasping* the proposition *Mary will come home* needn't be attended by my fearing that Mary will come home. For I could, at least in principle, grasp this proposition without *believing* it. (For me to entertain hypothetical propositions involving this proposition, I must obviously grasp this proposition; but my entertaining such hypothetical propositions doesn't involve my *believing* it. To entertain the hypothetical proposition 'if snow is green, then pigs fly' I must grasp the propositions *snow is green* and *pigs fly*; but obviously I can entertain the former hypothetical proposition without believing either that snow is green or that pigs fly.)

Is there any other proposition such that, merely by virtue of grasping it, I fear that Mary will come home? It doesn't seem so. For I can *consider*, and thus grasp, virtually any proposition without believing that there is the slightest chance of its being true and thus without having any fear in connection with it; I can, for example, consider the proposition *Mary will come home and torture me* and firmly believe that Mary will not come home. So I can consider, and thus grasp, virtually any proposition without feeling any fear of anything in connection with it.

Is there some *object* such that, merely by virtue of that object's being represented to me, I fear that Mary will come home? First of all, it seems to me that objects can't be represented to people without propositions being represented to them. An object cannot *just* be represented to me; it must be represented as having this or that characteristic – so the proposition that it has this or that characteristic must be represented to me if the object is to be represented to me. But *even if this isn't* true – even if objects can be represented to someone without propositions being represented to that person – this still doesn't help the representationalist. For there is no object O such that, merely by virtue of grasping O – by virtue of O's being represented to me – I have any propositional attitude (besides, perhaps, that of merely considering O). For what object is it the case that, merely by virtue of my grasping O, I fear that Mary will come home? None; for I can *think about* any object without fearing anything. I can think about a grinning Mary holding a blood-encrusted chainsaw without fearing anything. (Of course, my psychology might be such that, if I so much as contemplate such a horrifying object, I automatically feel some kind of dread. But surely this has to do with the peculiarities of my psyche. It is not *logically or metaphysically necessary* that I feel such dread upon thinking about this kind of object.) In general, I can think about any object without having any propositional attitude (other than, perhaps, that of merely considering the object).

Of course, what we've said about fearing that Mary will come home applies to propositional attitudes generally. There is no proposition P, no object O, such that merely by virtue of my grasping P or O I *doubt* that Mary will come home or *believe* that Mary will come home or *am sad* that Mary will come, and so on. Basically, there is something to propositional attitudes besides their having representational content; the attitudinal component of propositional attitudes is not in general assimilable to representational

content. Of course, the attitudinal component of a mental entity is a distinctly *mental* component of it. It follows that what is distinctly mental does not coincide with, and doesn't supervene on, what is representational: so representationalism is false.

Once again you have spoken a great deal about propositional content, but not about non-propositional content. I grant what you've said: there seems to be no *proposition* P such that *merely in virtue* of grasping P one has this as opposed to that propositional attitude. (And, I suppose, the same is true of any object that one might grasp.) So – as you said – the having of propositional attitudes isn't wholly assimilable to the mere apprehending of *propositions;* so (again, as you said) there is something to having propositional attitudes besides the mere apprehending of propositions (and objects): to have a propositional attitude is not *just* to be in a mental state with a certain propositional content. – Fair enough: you've made some case for *this.* But you haven't ruled out the possibility that to have a certain propositional attitude is *just* to apprehend some *non*-propositional representational content and that, consequently, the having of propositional attitudes *does* reduce to being in a mental condition that has a certain represnetaitonal (albeit non-propositional) content.

Fair enough; the objector has pointed out a lacuna in my argument. (Of course, if our earlier remarks about representational content are on the mark, they may to some extent vitiate what the objector is saying. But, admittedly, those remarks were much too contentious for me to use against the objector in the present context.) Let is fill this void. Suppose I *fear* that Mary will come home. Now is there *some* kind of representational content R such that

merely in virtue of my grasping R I have fear that Mary will come home? Is there *any* represnetaitonal content R – whether analogue or digital -- such that *merely* by virtue of my apprehending R I *ipso facto* am afraid that Mary will come home? Or suppose I (like Hilbert) believe that arithmetic is complete. Now is there some representational content – leaving it totally open whether R is a proposition, a non-proposition, a combination proposition/non-proposition – such that *merely* by virtue of my being in a mental state that encodes R (i.e. that is a grasping of R), I *ipso facto* believe that arithmetic is complete? Surely not. So the having of propositional attitudes very much seems *not* to be assimilable to the mere grasping of any kind of representational content. Conclusion: there is, at least in some cases, more to having a propositional attitude than being in a mental state that has a certain represnetaitonal content. (Now this leaves it open that in *some* cases the having of a propositional attitude reduces to the grasping of some representational content. It *might* be held – though I think it would be false to do so – that one cannot have a visual perception without *ipso facto* having some kind of belief: David Armstrong held this. Now a case might be made that to have a visual perception is *just* to apprehend a certain kind of content. Given these points, a case might be made– that, where visual perceptions are concerned, the having of a belief, and thus of a certain propositional attitude, supervenes on the mere *grasping* of a certain bit of representational content. Now I don't know whether this line of thought is correct or not. But it doesn't matter for our purposes. We have, I think, established that where *some* propositional attitudes are concerned, there is more to having those attitudes than *just* grasping a certain representational content; and, therefore, that propositional attitudes are not *in general* consistent with the claims of Representationalism.)

Importantly, we've seen in this section that *even if* intentionalism is

true, it doesn't follow that representationalism is true. Not all propositional attitudes have phenomenal content. For example, a person's believing that pi is transcendental has no phenomenal content, and neither does a person's *doubting* that pi is transcendental. We've seen that the difference in the attitudinal components of these propositional attitudes does not correspond to a difference in their representational components. At the same time, the difference in their attitudinal components cannot possibly correspond to a difference in their *phenomenal* components, since they have no such components. So someone's doubting that pi is transcendental is different in its distinctly mental properties from a person's believing that pi is transcendental; but this difference corresponds *neither* to a difference in phenomenal properties *nor* to a difference in representational properties. So even if intentionalism were correct –even if the phenomenal generally reduced to the representational – representationalism would *still* be false.

"You are arguing against a straw man," it will be said. "I can think of no one who thinks that the attitudinal differences between e.g. fearing that Bob is coming home and hoping that Bob will come home correspond to differences in the representational contents that they encode; I know of no one who holds that 'the attitudinal can be assimilated to the representational'."

It may be true that no one has ever maintained *specifically* that (as we have put it) the difference between fearing that Bob will come home and hoping that he will come home corresponds to a difference in the proposition encoded in these attitudes. Nonetheless, anyone who holds a strictly representationalist view of the mind – and there is no shortage of thinkers who (at least if we go by the letter of what they so) hold such a view – is, as I

have shown, *committed* to holding just this. (That they have failed to see that they were entangled in this commitment is to their discredit, not mine.)

V. We've seen already that, contrary to what intentionalism says, a difference in phenomenal content does not necessarily correspond to a difference in representational content. For we've seen what, by all appearances, are counter-examples to that thesis; and we have, if I am not mistaken, discredited the more plausible ways of trying to neutralize these counter-examples.

At the same time, the main premise of Byrne's argument for intentionalism seems to be true: if you focus on (say) your gustatory perception of the sweetness of the honey, and you imagine a change in the phenomenology of that perception, you are *ipso facto* imagining a change in what that perception represents. If you imagine a change in the phenomenology of your perception of the blue of the chip, you *ipso facto* imagine a change in what color the chip is represented to you as having. How is what we have seen – viz. that there are reasons for holding that intentionalism is false – to be reconciled with the apparent truth of this point of Byrne's?

One way to respond is by advocating a *restricted* intentionalism. Permit me to clarify this. We have noted, and made much of the fact, that if intentionalism is correct – i.e. if the phenomenal content of an experience is wholly determined by its representational content -- then for *any* pair of mental entities x and y (whether x and y are experiences or not), x and y must have the same phenomenal content if they have the same the same representational content. Of course, given what we've seen, it is very hard to believe that that this is the case. But, it might be argued, what *might* still hold is a form of intentionalism that is so weakened, so circumscribed, that it

entails only that for *some* pairs of mental entities x and y, x and y cannot differ in phenomenal content without differing in representational content. So, it might be thought, a *properly restricted* form of intentionalism might still hold: one that bears not on all pairs of mental entities, but only on pairs drawn from certain sub-classes of the mental.

The question then becomes: what are the appropriate restrictions? First of all, if what we've said is correct, then intentionalism must be restricted to *experiences.* As we've seen, a non-occurrent memory can have the same representational content as a perception. But clearly no perception has *any* phenomenal overlap with a non-occurrent memory or, indeed, with *any* mental entity that is not an experience. So the first restriction is this: a properly restricted intentionalism is not concerned with mental entities that are not experiences and that, consequently, have *no* phenomenal content (viz. 'dispositional' mental entities and also the deeply unconscious computational processes posited by Marr, Chomsky and other cognitive scientists).

The second restriction is not hard to identify. We've noted that perceptions belonging to different sensory modalities can overlap to a large degree in representational content, while having no phenomenal overlap; and we've also seen that this fact is inconsistent with (unrestricted) intentionalism. So a suitably restricted form of intentionalism is to hold *within* sensory modalities but not *between* them; a suitably restricted form of intentionalism will entail only that, given two perceptions *belonging to the same sensory modality*, those perceptions cannot differ phenomenally unless they differ representationally.

So, to sum up, our new, restricted form of intentionalism entails only that, if x and y are perceptions *belonging to the same sensory modality*, then x and y cannot differ phenomenally without differing representationally. Byrne does seem to have made a good case for *this* position.[317]

In a moment, I will try to show that even this restricted kind of intentionalism is probably not viable; more specifically, that *even within a single sensory modality* there can be phenomenal variation without any representational variation. But *even if we grant that this restricted form intentionalism is correct*, this would cast only a very limited light on the nature of the mental and, more specifically, on the nature of phenomenal content: indeed, it would say much less about the mental, and about the nature of phenomenal content specifically, than the fact that *un*restricted intentionalism is false. Our restricted intentionalism allows that there can be phenomenal variation without representational variation. (It allows that e.g. a non-occurrent memory and a perception can have the same representational content while having no phenomenal overlap.) So it straightforwardly concedes that phenomenal content is not, at least not in general, reducible to representational content. Consequently it leaves us in the dark as to what phenomenal content is. Indeed, supposing that the only correct form of intentionalism is restricted intentionalism, the only thing that fact teaches us is that phenomenal content is *not* the same thing as representational content (at least not in general)

Also our restricted form of intentionalism is quite incoherent. As just noted, our restricted form of intentionalism allows for phenomenal variation that does not correspond to representational variation. (It prohibits phenomenal variation without representational variation *only* where perceptions belonging to the same sensory modality are concerned.) Given this, let R be some bit of representational content that, in perception P1, is associated with phenomenal content C1. (Recall that the same bit of representational content can be 'associated with' different phenomenal contents in different perceptions. I can *see* that the ball is spherical and I can also *feel* that the ball is spherical. This representational content – viz. that the

ball is spherical -- is ‘associated with’ one bit of phenomenal content in my visual perception, and a very different one in my tactile perception.) Our restricted intentionalism allows for the possibility that some perception P2 also has R but does *not* have C (our restricted intentionalism allows this so long as P2 belongs to a sensory modality different from P1). Now if a perception P2 has R but does not have C, it follows that having R is not *sufficient* for having C. (If having R were sufficient for having C, then P2 would have C.) Since having R is not sufficient for having C, it follows, in turn, that *P1’s* having C is not *wholly* a function of its having R. (Since having R is not *sufficient* for having C, it cannot be that P1 has C wholly because it has R.) Now remember that (unrestricted) intentionalism is the view that the phenomenal content of an experience is a function wholly of its representational content. Now as we’ve just noted, our restricted intentionalism allows that any given bit of representational content can be associated with *different* phenomenal contents. So, it follows, our restricted form of intentionalism allows that *no* bit of representational content is *sufficient* for having any particular bit of phenomenal content. (If any bit of representational content were *sufficient* for having this or that kind of phenomenal content, then given *any* two mental entities sharing that representational content – regardless of whether those entities belonged to the same sensory modality or not; regardless, indeed, of whether they were both perceptions – that bit of representational content could not give rise to *different* phenomenal contents in the two entities. But, as we’ve just seen, our restricted intentionalism allows that one and the same bit of representational content can be associated with different phenomenal contents in different perceptions.) But if the latter is the case, it follows that a mental entity’s having a particular phenomenal content is *never* wholly a function of its having a particular representational content. So, to sum up, our restricted

intentionalism allows (by implication) that given *any* perception (indeed, given *any* mental entity), the phenomenal content of that perception (or entity) is *never* wholly a function of its representational content, But , of course, if the latter is the case, then *no* form of intentionalism – restricted or unrestricted – holds. For surely any form of 'intentionalism' worthy of the name holds that, in at least *some* cases, a mental entity has the phenomenal content it has wholly in virtue of having the representational content that it has: but it is just this, as we've just seen, that our restricted intentionalism implicitly denies. So our restricted 'intentionalism' really collapses into non-intentionalism.

However, even if we leave the contents of the last two paragraphs aside, it is *still* doubtful whether even a restricted intentionalism of the kind considered is correct: and this is so notwithstanding the strong *prima facie* case Byrne has made for it. (That said, I must to some extent disclaim what I am about to say: it is meant only to be suggestive, not probative. But I think that a rigorous refutation of even the restricted kind of intentionalism just considered could be coaxed out of the following considerations.)

The view to be defended is as follows. First of all, the relationship between our phenomenal contents and their physical causes is, presumably, a relatively *consistent* one; if our sensory surfaces are affected in the same manner on two different occasions, we don't experience radically different phenomenal contents as a result. So identity in physical input is (or tends to be) reflected in identity of phenomenal output. (At any rate, it is to be *hoped* that this is true. In so far as identity of physical input is followed by wild diversity of phenomenal output, then it seems our ability to obtain knowledge about the physical world is to that extent jeopardized. The viability of a representational system is predicated on, among other things, its *consistency*.) And, by the same reasoning, *similarity* of physical input is, presumably,

followed by similarity of phenomenal output. Consider the phenomenal output you have in response to your eardrums being disturbed by sound-wave oscillations of 440 hertz: let P1 be this phenomenal output. If your eardrums are disturbed by oscillations of 460 hertz, the resulting phenomenal output – call it P2 -- will be much closer to P1 than the phenomenal output resulting from noise of 700 hertz would be. In general there is a certain *commensurability* between them and their physical causes (in any case, this is what we must *suppose* if our phenomenal contents are to be of help in gaining knowledge of the outside world); the relationship between these two classes of events must be systematic – governed by some rule (a kind of 'rule of projection'). (Of course, there are different 'rules of projection' for different sensory modalities, since different modalities involve different types of phenomenal contents.)

Given all of this, suppose for the sake of argument that R is an arbitrary one of these rules. Here is what I suggest. If R is held constant – i.e. if the *overall* manner in which phenomenal contents (belonging to some one sensory modality) are paired off with their physical causes is held constant -- then any *particular* phenomenal shift will correspond to a representational shift. *However* if R is changed – if what is changed is the *overall* manner in which phenomenal contents are paired off with their physical causes – then a *particular* phenomenal shift need not correspond to a representational shift. Byrne, in effect, considers only cases of phenomenal shift that take place *within* a context where R has remained constant: thus he has made it appear as though phenomenal shift (a shift in the phenomenal content of one's experience) always corresponds to a representational shift (a shift in the representational content of one's experience). But when we consider cases where R itself is changed, then we see some reason to believe that a phenomenal shift does not always correspond to a representational shift.

These condensed and obscure remarks may make more sense if we further delineate the sound-example. Suppose that you have perfect pitch (so that whenever you hear a sound that is in fact, say, a middle C, you hear it *as* being a middle C – you don't have to consult some reference-point to figure out that what you are hearing is a middle C.) Given this, consider the kind of phenomenal content that you experience, when (under 'standard' acoustical conditions) a middle C (440 hertz) is played on a piano. Let p1 be this bit of phenomenal content. Now consider the kind of phenomenal content that you experience when (under standard conditions) a middle C# is played; let p2 be *this* phenomenal content. Let p3 be the kind of phenomenal content you have when a middle D is struck – and so on for all the other notes. Let **R1** be this rule. (So, just to be clear, **R1** is the rule that assigns p1 to a middle C, p2 to a middle C#, and so on.) Now if **R1** continues to hold – in other words, if your auditory phenomenal contents are paired off with sounds in accordance with **R1** -- then a shift in the phenomenal content of your auditory experience – say, a shift from having p1 to having p2 – will correspond to a shift in the representational content of your auditory experience (you will go from hearing a C to hearing a C#.) Up to this point we are in agreement with Byrne.

But now suppose that you have neurosurgery; and that, as a result, your 'pitch-spectrum' is altered. So now, whenever a middle C is played, you experience p2 (instead of p1); whenever a middle C# is played you experience p3 (instead of p2); and so on. So, in effect, the surgery would *replace* **R1** with a different rule – with a different pairing of phenomenal contents with disturbances of your eardrums.[318] (Let **R2** be this new 'rule of projection'.) Right after the operation, you would probably be thrown off by your phenomenal contents; when experiencing p2, you would think (wrongly) 'I am hearing a C#', when in fact you were hearing a D. But there

doesn't seem to be any reason why you couldn't adjust. Presumably you would, or at least could, become acclimatized to this shift; and after this acclimatization, when experiencing p2, you would automatically know (i.e. you would know without having to perform any conscious inference) that you were hearing a C: in effect, as a result of the said acclimatization, you would, in having p2, *hear* a C Thus, it seems to me --- and I admit that what I'm about to say is contentious – that, if you had such surgery, and if you became 'acclimatized' to the resulting shift in phenomenal contents, then the representational content associated with (say) p2 really would have shifted from a C# to a C (the same being true *mutatis mutandis* of p3, p4 and so on) So here, at least arguably, we have a case where, *within a single sensory modality*, there is a shift in phenomenal content that is *not* accompanied by a shift in representational content.[319] Before the surgery, p2 represented a C#; after the surgery, p2 represented a C. So because the overall rule of projection has changed – because the *overall* manner has changed in which phenomenal contents (of a certain sort) have been paired off with physical disturbances – a *particular* difference in phenomenal content doesn't correspond to a difference in representational content.

The basic idea here is this. It is not any bit of phenomenal content *per se* that corresponds to a bit of representational content. Rather it is the phenomenal content *as part of a larger pattern* that has representational content: a pattern of having certain phenomenal contents in response to certain stimuli. An analogy might help: any point in space can be represented by any number (actually, by any triple of numbers; but let us just speak of single numbers for the sake of expository ease). But if a given number is to represent a point in space, it must be part of some principled and consistent way of assigning numbers to points. It would be meaningless to assign a number to just *one* point: you wouldn't learn anything about the location of

that point from its number; and this is because there is no *intrinsic* relation between any number and any point in space. And it would also be meaningless to assign numbers to points in space if the arithmetical relations between the numbers did not have some kind of structural similarity to (some 'isomorphism' with) the geometric relations between the points thereby numbered. A given number meaningfully (informatively) labels a point in space iff it is part of a general assignment of numbers to points in space satisfying the following condition: for any two numbers x and y, if P(x) and P(y) are the points in space which x and y label, then one can infer the arithmetical relationship between x and y on the basis of the geometric relationship between P(x) and P(y) and *vice versa.* The idea is, roughly, that the 'rule of projection' – the rule pairing off numbers with points – must correspond to a kind of abstract, structural similarity holding between the two sets in question. And it is not anything *intrinsic* to a certain triple (say <45, 87, 44>) that makes it represent a particular point **P** in space: that triple is just a kind of place-holder; it represents **P** only because some rule of projection is countenanced, not because of anything intrinsic to that triple. The same arguably seems to be true of phenomenal contents. Given a particular bit of phenomenal content, it is not because of anything intrinsic to that content that it represent, say, a middle C; it is only because some rule of projection is operative. If a different one were operative, that same phenomenal content might represent a middle E. For a given phenomenal content to represent a middle C, or certain shade of pink, a certain temperature, it is necessary and sufficient that there be a certain abstract, structural similarity holding between your phenomenal contents and their respective physical causes. But *different* pairings of phenomenal contents and physical causes could, presumably, realize this abstract similarity: so a given phenomenal content could betoken a C or a C# or a D; and, consequently, a given representational

content could be associated with different phenomenal contents belonging to the same modality. So, if we are right, there is no absolute correspondence between phenomenal and representational content; and different phenomenal contents, even within the same modality, can bear the same representational content, contrary to what even the restricted intentionalist holds.

References

Brentano, Franz

[1] *Psychology From an Empirical Standpoint* edited by Linda L. McAlister with a new introduction by Peter Simons. Translated by D. B. Terrell, Antos C. Rancurello, and Linda L. McAlister. London and New York: Routledge, 1995.

Byrne, Alex

[1] 'Intentionalism Defended'

The Philosophical Review, Vol. 110, No. 2 April 2001

Dretske, Fred

[1] *Naturalizing the Mind* (The M.I.T. Press 1995) chapter 3, section 2

Harman, Gilbert

[1] 'The Intrinsic Qualities of Experience' in *Philosophical Perspectives 4*, 1990.

Heidegger, Martin

[1] *Being and Time,* transl. by J. Stambaugh, SUNY Press 1996

MacDowell, John

[1] *Mind and World* Harvard 1994

Nagel, Thomas

[1] 'What is it like to be a bat?' in *The Nature of Mind* edited by David Rosenthal Oxford 1990

Sartre, Jean-Paul

[1] *Being and Nothingness: A Phenomenological Essay on Ontology*, tr. by Hazel E. Barnes (Washington Square, 1993)

Review of Erin McReady's review of Conceptual Atomism and the Computational Theory of Mind

Introduction Erin McReady, a distinguished linguist, recently published a review of my (2007) book *Conceptual Atomism and the Computational Theory of Mind: A Defense of Content Internalism and Semantic Externalism* (CAC). While I do not question Dr. McCready's excellence as a scholar, I submit that he did not do himself justice in this review of his, which consists of shoddy arguments directed against positions that McCready either mischaracterizes or that I simply didn't advocate. For example, McCready says that my book is an attempt to refute semantic externalism, even though the book is subtitled: *A Defense of Semantic Externalism and Content Internalism.* This was not a slip of the pen on McCready's part: his criticisms of my book are hewed to the assumption that semantic externalism is, indeed, one of my targets.

In this 'counterreview,' as it were, I will summarize what I said correct those distortions by outlining some of the main contentions of that book. Whenever we put forth a view with which McCready took issue, we will pause to determine the merit's of McCready's position.

In Part 1, we'll make a case for semantic externalism and against content internalism. In Part 2, we'll make a case against CTM, having first shown that, given the falsity of content externalism, it loses much of its credibility. In Part 3, we'll put forth and defend an answer to the question "what are propositions?," which, if correct, provides further support for the anti-CTM points to be made in Part 3 and which also, unlike rival analyses, establishes a natural and data-consistent between the structures of propositions, on the one hand, and, on the other, those of the sentence(-tokens) that express them. In Part 4, we'll address, and refute, McCready's superficially plausible

contention that there are *bona fide* sentences that are utterly meaningless, in the process of doing which we'll find possible ways of dissolving some semantic and logical antinomies that have bedeviled foundationalist programs in mathematics.

Part 1 **1.0 <u>Is linguistic meaning a strictly psychological notion?</u>** In *Naming and Necessity*, Kripke decisively showed that proper names are not definite descriptions. In doing this, Kripke was making a specific point concerning the semantics of a specific class of expressions. It might therefore seem that what Kripke was doing in NN, though non-trivial, was too narrow to be of great philosophical importance.

But implicit in what Kripke said about proper names are many deep and important principles concerning the very nature of meaning and the relationship between language and thought. In 1975, Hilary Putnam published an article titled *The meaning of 'meaning'* (henceforth, 'MM') in which he explicitly stated one of these principles and cogently argued for it. What he showed is that what we think does not, by itself, establish what our words and sentences mean.

It's obvious that what we think *largely* determines what it is that our expressions mean. The word 'snow' could refer to anything; it would be easy to create a language in which it referred, not to so snow, but Julius Caesar. (Let's create such a language right now. Let L be a language that is just like English except that, in L, 'snow' refers to Julius Caesar. L is a perfectly good language – it's as good as a language as English.) It's obviously at least partly because of what some people were thinking at certain times that 'snow' refers to snow, and not to Julius Caesar.

But Putnam made it clear that there is an important *non*-psychological

dimension to linguistic meaning. Given only that, any given moment, person X is psychologically and behaviorally indistinguishable from person Y, it does *not* follow that the words leaving X's mouth have the same meanings as the words leaving Y's mouth. So, supposing that at any given moment X is having a given thought if and only if Y is having an exactly similar thought or feeling, and that X is producing a given utterance if and only if Y is producing an exactly similar utterance, it does *not* follow that what is meant by X's utterance coincides with what is meant by Y's. Even though the psychological state that gave rise to the one utterance is just like the psychological state that gave rise to the other, the meaning of the one may diverge from that of the other.

1.1 <u>Putnam's argument</u> Let Twin-Earth be a planet in some distant galaxy that satisfies the following conditions. Given any person x on Earth, there is a person y in W who is an atom-for-atom duplicate of x. At any given moment, y has a given thought if and only if x has that thought; y has a given feeling if and only if x has that very feeling. y acts a certain way if and only if x acts in an exactly similar way; y makes a certain noise if and only if x makes an exactly similar noise. Thus, Earth people are psychologically and behaviorally exactly like Twin-Earth people. They think, feel, do, and say the very same things at the very same times.

It is merely coincidence that Earth people are in lock-step with Twin-Earth people. There is no causal interaction between Earth-people and Twin-Earth people. No messages are sent from the one planet to the other; nobody on either planet in any way influences anyone on the other.

Even though Earth and Twin-Earth are extremely similar, there is an important difference between them. On Twin-Earth, the liquid in bathtubs, drinking-glasses, lakes, etc. isn't H20 – it's XYZ. ('XYZ' is an abbreviation

for some long chemical formula.) With that qualification, W is, at any given moment, exactly like Earth.

Let's continue our story. Smith is a psychologically normal, English speaking, adult Earthling who lives in the 1600's. One day, after doing back-breaking work for several hours, Smith is desperately wants to drink some cool water. Addressing his companion, Smith says:

(SE[320]) 'When people are thirsty, they find it more refreshing to drink cool, glacial water than warm, swamp-water; if there were no clean water, people couldn't survive .'

This utterance takes place at time t.

Important point: SE doesn't refer to a sentence. It refers to a specific *utterance* of a sentence. It refers to the specific burst of noise produced by Smith at time t. (Different people can *utter* a given sentence; different people can *utter* the sentence uttered by Smith (viz. 'when people are thirst, they find it more refreshing to drink cool, glacial water than…')

Bear in mind here that, in Smith's time, nothing was known about the chemical composition of water. People obviously had thoughts about water; they wanted to drink it; bathe in it, and so on. But they didn't know that it was H2O; they didn't even grasp the concepts needed to grasp the concept of an H2O-molecule.

On Twin-Earth, there is somebody who, at any given moment, is psychologically and behaviorally just like Smith. (This follows from what we said earlier.) Let Twin-Smith be that person. At any given time, Twin-Smith makes the very noises as Smith and, in so doing, expresses the very sentiments. (This too follows from our earlier points.) Thus, at time t, Twin-Smith says:

(TS[321]) ‘When people are thirsty, they find it more refreshing to drink cool, glacial water than warm, swamp-water; if there were no clean water, people couldn’t survive.’

Important point: TS doesn’t refer to a sentence. It refers to a specific *utterance* of a sentence -- to the specific burst of noise produced by Twin-Smith at time t.

Question: Does SE have the same meaning as TS? I am not asking whether Smith and Twin-Smith coincide in respect of what *they* mean by their respective utterances, but whether SE and TS coincide in respect of what *they* mean. The question doesn’t concern what Smith and Twin-Smith are trying to convey through their respective utterances; it concerns what those utterances themselves mean.[322]

SE and TS *don’t* have the same meaning. They have different propositions for their respective meanings, and the one proposition is true in circumstances where the other is false, and false in circumstances where the other is true.

‘Water’ refers to the substance whose chemical composition is H2O. That’s a simple fact. A thoroughly competent speaker of English can easily fail to *know* that ‘water’ refers that substance. Newton didn’t know it, even though he was an exceptionally intelligent person who spoke perfect English. But that doesn’t matter. Here on earth, there is a certain liquid x such that x is found in bathtubs, drinking glasses, lakes and oceans, and such that that is referred to as ‘water.’ That substance was referred to as ‘water’ for many hundreds of years before anything was known about chemistry or, in particular, about the chemical composition of that substance. When, finally,

that substance was analyzed, it turned out to consist of H2O molecules. Thus,

(WAT1) there is some substance x such that x is found in bathtubs, drinking glasses, lakes and oceans, and such that x consists of H2O molecules, and such that: ‘water’ refers to x.

Of course, we now know that the substance referred to as ‘water’ is H2O. And if we were to discover some substance that consisted of XYZ, not H2O, but was otherwise just like water, at least as far as the conditions that are likely on the surface of the Earth are concerned, that other substance wouldn’t be H2O. The chemists who discovered XYZ would *not* come conclusion that water is *sometimes* H2O and is sometimes XYZ. No – they’d say that water is *always* H2O, but that a different substance, namely, XYZ is extremely water-like, at least within certain horizons.

When they weren’t referring to XYZ by its chemical name, chemists would *not* refer to it as ‘water.’ Maybe they’d refer to it pseudo-water’ or as ‘fool’s water’ – but not ‘water.’ This practice of theirs wouldn’t be a case of hair-splitting pedantry. If chemists *did* refer to this other substance as ‘water’, their doing so would violate beliefs that are deeply embedded in commonsense. Even the most philistine, intellectually narrow human being desperately wants real diamonds, not Zirconium imitation-diamonds, and thus knows that being diamond-*like* isn’t the same thing as being a diamond. Similarly, everybody knows that being water-*like* isn’t the same thing as being real water. And even if some people don’t know it, they’d quickly learn it, no matter how scientifically uninformed they were, if the distinction between water and XYZ became a financially significant one, like the distinction between diamonds and chunks of zirconium paste resemble diamonds.

For the same reasons *mutatis mutandis*, on Twin-Earth 'water' would refer to XYZ, not H2O. And it would *always* have referred to XYZ – even before scientists on their planet discovered its chemical composition.

Before proceeding, let's deal with an objection that is likely to be heard:

An English-speaking physical chemist obviously knows a lot about water that other English-speakers don't. But what the physical chemist that other English speakers don't has nothing to do with the English language or, therefore, with *semantics*. Given only that somebody doesn't have a PhD in physical chemistry, it doesn't follow that they have any *linguistic* deficits – that they don't *really* know the identities of the semantic rules governing expressions like 'water', 'gold', 'lead', etc. This is all quite uncontroversial.

But it's inconsistent the view you're advocating. You say that, for something to be water, it isn't enough that it have certain macro-characteristics: it must have specific chemical composition – it must be H2O. Given that, as a matter of semantics, 'water' refers to water, what you're saying entails that, as a matter of semantics, 'water' refers to H2O. But if that were right, one couldn't master the English without knowing the chemical composition of water or, by obvious extensions of this argument, the chemical composition of gold, aluminum. But that's preposterous.

One doesn't have to know anything about chemistry to have a complete mastery of English (or any other natural language), and it would be preposterous to deny that. But the view I'm describing doesn't have that consequence. There is some individual x such that the semantic rule for the expression 'Barack Obama' is to the effect that:

(BO) ‘Barack Obama’ refers to x.

BO pairs off a name with an individual. If one doesn’t know which individual that is, then one doesn’t know the semantic rule for ‘Barack Obama’ and, therefore, one doesn’t know all of the semantic rules constitutive of the English language in 2009.

But given any entity, there are many ways of knowing which thing it is. There are infinitely many ways of singling out any given entity. This is a consequence of the fact that, as a matter of logic, any given object satisfies each of infinitely many different descriptions. Let P be the region of space you were occupying the instant you began reading this sentence, and let t be that instant. You uniquely satisfy each of the following predicates: ‘thing that occupied p one minute before time t+one minute’, ‘thing that occupied p two minutes before time t+two minutes’, and so on. We’ll henceforth abbreviate the first of these predicates as ‘P1’, the second as ‘P2’, etc.

For argument’s sake, let’s suppose that your name is ‘Jack Jones.’ There is some x such that x uniquely satisfies each of P1, P2, P3, etc. and such that, in order to the know semantic rule governing that expression, one must know that:

(JJ) ‘Jack Jones’ refers to x.

If Smith knows that:

(JJ75) there is some x such that x uniquely satisfies P75 and such that: ‘Jack Jones’ refers to x

and Jones knows that:

(JJ143) there is some x such that x uniquely satisfies P143 and such that: ‘Jack Jones’ refers to x,

Smith and Jones are to that extent in complete semantic accord. There is some one x such that x satisfies each of P75 and P143, and such that, given any property psi, when Smith says

(JJ1) ‘Jack Jones has psi’

he is affirming that

(JJ2) x has psi,

and such that when Jones utters JJ1, he too is affirming that x has psi. Further, it is each individual’s *intention* in uttering JJ1 to affirm JJ2. Each wishes to do is to attribute psi to some person – nothing more. Neither wants to say anything about his way of thinking about that person. Neither wishes to say of P75 or P143 or any other property that it is uniquely instantiated or that all its instances are psi’s.

Points exactly similar to those just made about ‘Jack Jones’ hold with respect to ‘water.’ At time t, bathtub B is filled with pure water, on which your rubber ducky is floating. Your mother points to that B and says:

(*) The liquid in the tub is known as ‘water.’

A few days later, at time t*, you burn your finger in scalding water. Given

that water is H2O, it follows that each of the following statements is true:

(1) there is some liquid x such that, at time t, x is what your rubber ducky is floating on; moreover, ‘water’ refers to x. (2) there is some liquid x such that, at time t, x is what your bathtub is full of; moreover, ‘water’ refers to x.
(3) there is some liquid x such that x and x alone is H2O, and such that ‘water’ refers to x
(4) there is some liquid x such that x and x alone is H2O, and such that ‘water’ refers to x.
(5) there is some liquid x such that , at time t, x and x alone is what scalded your finger; moreover, ‘water’ refers to x.
(6) there is some liquid x such that x and x alone is H2O and such that, at time t, x and x alone is what scalded your finger; moreover, ‘water’ refers to x.

And so on.

To know what ‘water’ refers to, it is necessary only that you know *some* truth of the form:

(n) there is some substance x such that x uniquely has property phi; ‘water’ refers to x.

Since water uniquely has each of infinitely many different properties, (n) is true for each of infinitely many values of phi. Where any given one of these truths is concerned, knowing it is sufficient, but not necessary, for knowing what ‘water’ refers to and, therefore, for knowing the semantics of that expression.

Since (n) is a sentence-schema, not a sentence proper, it does not itself

express any one of these truths. But in what follows, take (n) to stand for an arbitrary one of them.

Only the underlined part of n describes a semantic rule. The rest describes the information through which one grasps that rule. Anyone who knows the semantics of water knows this (though they might recognize what it is that they know in the statements just made).

Let's tell another story about Smith and Jones. Both are competent speakers of English. Smith knows that:

(7) there is some liquid x such that x is what is in bathtubs and drinking-glasses; moreover, 'water' refers to x.

Jones, who is more scientifically minded than Smith, knows that:

(8) there is some liquid x such that x is H2O and such that creatures with gills can breathe when immersed in x and creatures with lungs cannot; moreover, 'water' refers to x.

There is some liquid x such that individual's sole intention in saying

(w2) 'water covers most of the Earth's surface'

is to say:

(w3) x covers most of the Earth's surface.

Neither wishes to affirm either that

(w4) there is some liquid x such that x is what is in bathtubs and drinking-glasses, and such that ‘water’ refers to x; moreover, x covers most of the Earth’s surface,

or that

(w5) there is some liquid x such that x is H2O and such that creatures with gills can breathe when immersed in x and creatures with lungs cannot; moreover, x covers most of the Earth’s surface.

Each individual’s intention in uttering w2 is to ascribe a certain property to a certain substance. They have the same substance in mind. They wish to attribute the same property to it. Neither wants to say anything about how he thinks about that substance. Neither wants to attribute any property to anything other than that substance. Thus, what Smith means when he says ‘water covers most of the Earth’s surface’ is exactly what Jones means when *he* utters the same sentence. And so long as each knows that the other agrees with him about what it is ‘water’ that refers to, any difference between them in respect of how they think about its referent will in no way jeopardize the one person’s ability to understand the other. To put it more formally: provided that, for some x, each believes that ‘water’ refers to x and each knows that the other believes this, the differences between these to individuals in respect of how they think about this substance won’t jeopardize their ability to communicate with each other.

1.2 Putnam’s semantic externalism (continued) These points make it clear why it is that Smith and Twin-Smith, though driven to make the very same noises by the very same internal conditions, are nonetheless affirming

different propositions. Let's suppose that when Smith was a child, his mother pointed to the liquid in the tub in which Smith regularly bathed, which Smith recognized as the liquid that he had to drink in order to live, and said: 'That is water.' The message that is communicated to Smith is along the lines of:

(8) There is some liquid x such that x is what I bathe in and must also drink in order to survive and 'water' refers to x.

Given that liquid indicated by Smith's mother is, of its very nature, H2O, it follows that when, at time t, Smith utters SE, his utterance is true exactly if:

(SE*) 'When people are thirsty, they find it more refreshing to drink cool, glacial H2O than warm, swamp-H2O; and if there were no clean H2O, people couldn't survive

Because he lives hundreds of years before the birth of modern chemistry known, Smith has no idea what it would even mean to identify water with H2O. Moreover, *nobody* who is alive at the time has any such idea. Nonetheless, the statement he is making does concern H2O. This shows that:

(PTM) the meanings borne by our utterances are not determined solely by the thoughts we have; there is a strictly non-psychological, strictly environmental dimension to linguistic meaning.

Given what we know about Twin-Smith and Twin-Earth, it follows that, when Twin-Smith utters TS, what is saying is true exactly if:

(TS*) 'When people are thirsty, they find it more refreshing to drink cool,

glacial XYZ than warm, swamp-water; if there were no clean XYZ, people couldn't survive.'

And this is obviously consistent with PTM.

The doctrine that PTM is true known as *semantic externalism.* It is so-called because, according to PTM, the determinants of the semantics of our expressions are (in some cases, in some respects) external to us.

2.0 Burge's blunder – the (mis)step from semantic to content externalism

PTM is a deep, important, and original insight, and deserves sole credit for it. And unlike many of the philosophers who were influenced by his important article, Putnam's views as to what PTM does and does not imply were correct. In MM, Putnam says that Smith's *words* don't have quite the same meanings as Twin-Smith's *words*, even though Smith and Twin-Smith have the same *thoughts.*

Under pressure from other philosophers, especially Tyler Burge[323], Putnam later rejected this correct and profound insight, and he came to accept Burge's position, which is that, not only do Smith's *words* have different meanings from Twin-Smith's, Smith's *thoughts* don't have the same contents as Twin-Smith's.

Burge's work has been enormously influential. It didn't just influence Putnam, but analytic philosophy as a whole. Thanks largely to Burge's work, it came to be widely held for twenty years or so that Smith and Twin-Smith have different *thoughts*. This must be understood aright. Nobody denies that Smith and Twin-Smith have numerically different thoughts. Burge is making the much less innocuous claim that what Smith is thinking is different from what Twin-Smith is thinking. In Burge's view, just my thought that it's hot

during the summer has a different content from your thought that it's cold during the winter, so, according to Burge, the thought that Smith expresses in saying, for property psi, 'water has psi' is different from the thought that Twin-Smith expresses by means of that same utterance.

This interesting chapter in the history of analytic philosophy began in 1979, four years after Putnam published MM, with the publication of Burge's article 'Individualism and the mental' (henceforth, 'BI'[324]). Here is what Burge argues. There is a planet, which we'll call 'Twin-Earth', that is exactly like Earth, with one qualification to be stated in a moment. The events on Twin-Earth are in lockstep with those on Earth. So given that earthling Smith sneezes at a certain time, Twin-earthling Twin-Smith sneezes in the exact same way at the exact same time.

Given what we've said, it follows that English is spoken by many people on Twin-Earth. Twin-Smith is one of these people. On our planet – on Earth, in other words -- 'arthritis' refers to a disease exclusively of the joints. On Twin-Earth, 'arthritis' refers to a malady not *only* of the joints, but also of certain bones (as opposed to the cartilaginous joints linking bones to one another). So, on Twin-Earth, 'arthritis' doesn't refer to the same disease that it refers to on Earth. It refers to a similar, but distinct ailment. Let us refer to that ailment as 'twin-arthritis.'

On Earth, Smith regards the sentence:

(AU[325]) 'arthritis is unpleasant'

as expressing a truth. And because he personally suffers from arthritis, he often utters AU. (In this context, it is to be assumed that Smith is speaking sincerely, is cognitively competent, etc., and these same assumptions are to

be made about Twin-Smith).

Like Smith, Twin-Smith regards AU as expressing a truth. But, according to Burge, the thought underlying Twin-Smith's utterance of AU is different from the thought underlying Smith's utterance of it. Twin-Smith is thinking about twin-arthritis, not arthritis, and Smith is thinking about arthritis, not twin-arthritis. In a moment, we'll evaluate this important claim.

Burge concludes from this that two people who are atom-for-atom duplicates of each other, each a perfect physiological and psychological clone of the other, may have thoughts with different contents. Twin-Smith is thinking *Descartes* (or Twin-Descartes, rather) *was a great philosopher*, whereas Smith is thinking *Ted Danson is a great philosopher*, notwithstanding that Smith and Twin-Smith are, by Burge's hypothesis, psychologically exactly the same.

Is Burge right? First of all, he's right that, on Twin-Earth, 'arthritis' does *not* refer to arthritis, and instead refers to Twin-arthritis. This is precisely what Putnam showed. But is Burge right that Smith and Twin-Smith are *cognitively* different?

No. How does Smith learn what 'arthritis' refers to? Here's a possible answer (and it's representative, in the relevant respects, of all the other possible answers). Addressing Smith, so and so says:

(A1) 'the unpleasant inflammation in your knees is due to arthritis; arthritis can do terrible damage to one's joints.'

A1 is to the effect that:

(A2) there is some x such that the swelling in your knees is due to x and such that x can do terrible damage to one's joints; moreover, 'arthritis' refers to x.

For now familiar reasons, it follows that when Smith hears someone say:

(A3) ‘arthritis has psi’,

the message that is communicated (not literally meant) is:

(A4) There is some x such that the swelling in your knees is due to x, such that x can do terrible damage to one’s joints, and such that ‘arthritis’ refers to x; moreover, ‘arthritis’ refers to x.

Remember that Twin-Smith’s life precisely parallels Smith’s. It therefore follows that, when Twin-Smith hears someone utter (A3), (A4) is the message that is communicated (not literally meant) is (A4).

Thus, contrary to what Burge says, Smith’s thoughts *do* have the same content as Twin-Smith’s. It’s true that, for the reasons given by Putnam, the *sentences* they use to express those thoughts don’t always have the same literal meanings. But unless one countenances the obvious false assumption that what we think coincides exactly with what our expressions mean, it doesn’t follow that Smith’s thoughts have the same contents as Twin-Smith’s.

2.1 <u>Literal meaning vs. cognitive content</u> An utterance on Earth of ‘arthritis is painful’ *does* have a different literal meaning from such an utterance on Twin-Earth. There is some malady x, namely arthritis, such that x only afflicts the joints, such that an utterance *on Earth* of

(A5) ‘arthritis is painful’

is true just in case:

(A6) x is painful,

it being irrelevant whether some disease that afflicts things *other* than the joints is also painful. And there is some y, namely twin-arthritis, such that y does not afflict only the joints and such that an utterance on *Twin*-Earth of

(A5) 'arthritis is painful'

is true just in case:

(A6) y is painful,

it being irrelevant whether some disease that afflicts *only* the joints is also painful.

But, contrary to what Burge assumes, this doesn't mean that

2.2 A second Burge-style argument [326] Let W1 be the actual world and let W2 be some other possible world. In each of W1 and W2, my friend Max has an identical twin Gary. Gary and Max are cosmetically, physiologically, and behaviorally completely and totally indistinguishable. To know that one is seeing Max as opposed to Gary, or *vice versa*, one has to rely on ulterior knowledge (e.g. that Gary is now in Finland and that he couldn't possibly be the person with whom one is now lunching in New York).

In W1, at time t, my friend Sally and I are ten feet away from Max, and I'm looking right at him. But in W2, at time t, Sally are ten feet away from

Max's identical twin, Gary. Otherwise, W1 at t and W2 at t are indistinguishable (at least at time t). Given only the information available to me by virtue of having the perception involved in my seeing Max, it is not a even a theoretical possibility that I might encounter some evidence or have some experience on the basis of which I could rationally conclude I was in the one world as opposed to the other.

Moreover, the experiences that I have in W1 had up until time t have been qualitatively identical with those that I've had in W2 up until time t. So neither my current experiences nor my memories nor the two combined could possibly enable me to determine whether I'm in the one world as opposed to the other.

It is to be assumed in what follows that all my utterances are sincere – they express beliefs that I actually have.

In W1, while pointing at Max, I tell Sally:

(U1[327]) 'That entity is wearing a plaid shirt.'

(In due course, it'll be clear why I'm using the word 'entity', as opposed to individual.) In W2, while pointing at Gary, I tell Sally:

(U2) 'That entity is wearing a plaid shirt.'

U1 and U2 are specific *utterances*. So U1 is an utterance that occurs in Max's presence in W1, and U2 is an utterance that occurs in Gary's presence in W2. In each world, I know that I'm seeing *either* Gary *or* Max, but in neither world do I know which of the two I'm seeing.

In W1, there is some individual x, namely x, such that such x^1Gary and such that U1 is true exactly if:

(MP[328]) x is wearing a plaid-shirt.

In W2, there is some other individual y, namely Gary, such y¹Max and such that U2 is true exactly if:

(GP) y is wearing a plaid-shirt.

MP is true just in case *Max* is wearing a plaid-shirt, it being irrelevant whether Gary is wearing one. And GP is true just in case *Gary* is wearing a plaid-shirt, it being irrelevant whether Max is wearing one. Neither entails the other; neither entails the negation of the other. MP and GP are different propositions.

U1 expresses what I'm thinking in W1, and U2 expresses what I'm thinking U2. U1's truth-conditions coincide of the thought I'm having in W1, and are therefore given by MP; and U2's truth-conditions coincide of the thought I'm having in W2 and are therefore given by GP. Thus, the belief expressed by U1 has a different content from the belief expressed by U2.

Of course, in W1, my visual experience results from light bouncing off Max, not Gary; and in W2, it results from light bouncing off of Gary, not Max. But apart from that, there is absolutely nothing that at that time distinguishes my situation in the one world at time t from my situation in the other world . My intracranial situation in W1 is just like your intracranial situation in W2.

But we just saw that the content of what I'm thinking in W1 at t is different from the content of what I'm thinking in W2. This difference must derive from the fact that, in W1, my mental state has a certain external cause (light bouncing off of Max), whereas in W2 it has a different external cause

(light bouncing off of Gary).

2.3 Burge's failure to distinguish what we know directly from perception from what we infer from perception

The conclusion of this argument is false, and there is thus something wrong with the argument itself. The problem with it is that it assumes that the relationship between linguistic meaning and cognitive content is much tighter than it actually is.

Let's suppose the following. I'm in some world W3 (not identical with W1 or with W2). I'm with my friend Sally. I get her ten dollars that, sometime this day, we'd see somebody wearing a plaid shirt. She doesn't know, or know of, Max or Gary. W3 is exactly like each of W1 and W2 *except* that, in W3, the thing standing in front of me isn't Max or Gary: it's a robot that looks exactly like both of them. I don't (yet) know that such a robots exists. I'm friends with Max and Gary. I happen to know that Gary is out of the country. I also know that Max is somewhere in the vicinity. So, given that I'm looking at something that looks exactly like Max, I believe that I *am* seeing Max. I'm with my friend Sally. I see that that the thing I'm beholding is wearing a plaid shirt, and I say:

(U3) 'That entity is wearing a plaid shirt.'

Let VR[329] be the visual perception that in which my seeing this thing consists. VR is utterly high-resolution and veridical. I'm seeing the thing before me, whatever it is, *as* it is. Given that this thing looks exactly like Max and Gary, my seeing it as it is involves my having a perception exactly like the one that I'd have in that very situation if *ceteris paribus*, it was Max or

Gary I was seeing.

Thus, VR does not itself tell me who it is that I'm seeing. Given only what VR tells me, that thing could be Max or Gary or some third person (or thing) . VR does tell me that I'm seeing somebody with such and such characteristics, among them that of wearing a plaid shirt.

I'm wrong about who that entity is. But that doesn't affect the belief I'm trying to get across. I don't care who or what that entity is, and neither does Sally. All we care about is whether it's wearing a plaid-shirt. I correctly believe that it is wearing a plaid shirt, and that's the belief I'm trying to get across by uttering U3. Therefore, the belief embodied in U3 is correct.

This has two very important consequences, each of which is incompatible with Burge's position.

Consequence #1: There is no particular x such that VR tells me that it's x I am seeing. This follows from the fact, just noted, that VR does not itself apprise me of the identity of the thing I'm seeing.

Consequence #2: There is no particular x such that the belief embodied in U3 is correctly only if *x* is plaid shirt. This follows from the fact, earlier noted, that the belief embodied in U3 is correct as long as the thing we're seeing is wearing a plaid shirt, it being irrelevant who or what thing is.

It is a fact that Max is the person I'm seeing -- that he's the object of VR (or one of them). But given only what VR is telling me, I could be seeing any number of different individuals. What VR tells me is that there is *somebody* having such and such characteristics – that there is *somebody* standing in a certain place and wearing a plaid shirt. So the content of VR is existential. It tells me that:

(CVR[330]) There is some entity x such that x has such and such characteristics, such that x is standing is in a certain place and such that x is

wearing a plaid shirt.

Obviously, CVR is at least one of the things that, on the basis of VR, I believe to be true. And it's the relevant belief in this context, since it's the one embodied in U3.

Supposing that VM is the perception I'm having in U1 of Max, and that VG is the perception I'm having in U2 of Gary, it follows that the belief embodied in U3 has the very same content as the belief embodied in U1 as well as the belief embodied in U2. Remember that, from an experiential standpoint, no two of these situations are distinguishable, even in principle.

Let's state as clearly and explicitly as possible how Burge goes wrong. Burge says that what VR is telling me is *different* from what VM is telling me and also from what VG is telling me. Burge thinks that there is some x, namely a certain robot, such that VR is correct just in case:

(bx) x is wearing a plaid shirt.

And Burge thinks that there is some y, namely Gary, such that VG is correct just in case:

(by) y is wearing a plaid shirt.

Finally, Burge thinks that there is some z, namely Max such that VM is correct just in case:

(bz) z is wearing a plaid shirt.

Of course, none of these three propositions is equivalent with other; none so

much as entails either of the others. So if Burge is right, then the proposition encoded in my visual perception in W3 is different from my visual perceptions in W1 and W2, and the propositions encoded in the latter two perceptions are different from each other.

But Burge is not right. Given only what VR is telling me – in other words, given only what I know on the basis of VR *without* the help of pre-existing, background knowledge derived from some source other than VR – I can't possibly know that I'm seeing robot x as opposed to person y as opposed to person z as opposed to Martian w. the content of my perception on W3 is, quite obviously, neutral with respect to the identity of the object of that perception, and its content is therefore, in that respect, general and, more specifically, is given by CVR.

Russell (1927) makes this point very clearly.

"Prejudice tells you that you see the *same* [Russell's emphasis] table on two different occasions; you *think* that experience tells you this. If it really were experience, you could not be mistaken; yet a similar table may be substituted without altering the experience…there is nothing to show that one identical entity causes the two sensations." [331]

The message embodied in VR goes through regardless of who it is that I am seeing to be wearing a plaid shirt. That identity-independent message is embodied in the belief underlying in my utterance of 'that guy is wearing a plaid shirt', and it's embodied in that utterance. Therefore, Burge is simply wrong to identify the content of that belief with bx. For exactly similar reasons, Burge is just plain wrong to identify bx with what VM tells my and with the belief embodied in U1, and he's just plain wrong g to identify by

with what VG tells my and with the belief embodied in U1.

Of course, there are differences between those situations. In one of them, it's Max who's standing in front of me; in the second, it's Gary; in the third it's a robot. But these differences don't lead to any experiential differences, since those three things all look exactly the same. In none of these worlds can the content of my visual perception, taken by itself, register that it's Max, Gary, or some third thing that's in front of me. Given that CVR is the content of my visual perception in W3, it follows that CVR is also the content of my perception in U1 and also of my perception in U2. (The content of my visual perception in the one world coincides with the content of my visual perception in each other.) Given that, in W3, it's my belief that CVR is true that leads me to utter U3, it follows that it is that same belief that prompts to utter U1 and also U2.

2.4 <u>Burge's failure to distinguish literal meaning from cognitive content</u>

U3 does indeed have different truth-conditions from each of U1 and U2. But it doesn't follow that the thought underlying my producing U3 is different from either of the thoughts that prompted my producing those other two utterances.

Not only does it not follow: it isn't true. What VR is telling me, and what I believe strictly on the basis of it, is CVR. It's true that I have other beliefs. I believe (incorrectly) that it's Max I'm seeing. But that belief is a derivative of my belief that CVR is true. And the belief that prompts to utter U3 – the belief I'm expressing – is true, no matter who or what the plaid-shirt wearer turns out to be.

An extension of our story about W3 makes this clear. In plain view of Sally and myself, this entity takes its face-plate off, revealing a complex web of circuitry, as opposed to various organic tissues. I now know that it's a

robot, and, given my pre-existing knowledge that Max isn't a robot, that it therefore isn't Max. Now that I know that it was a robot that I was seeing, am I to conclude that I the belief embodied in U3 is wrong? No. If I look at what is in fact a green car and say 'that car is green', what I'm saying is correct and the belief embodied in that utterance is correct, even I also happen to have the false belief the car belongs to Smith when, in fact, it belongs to Jones. For exactly similar reasons, the belief embodied in U3 is correct.

Nor am I conclude that VR wasn't veridical. VR was high-resolution and veridical. The problem had to do with the fact that, in having VR, the thing I was accurately seeing, though not identical with Max, looked just like him, and with the consequent fact that, on the basis of my correct belief in CVR, I erroneously assumed that it was Max I was seeing.

2.4.1 <u>Burge's failure to distinguish literal meaning from cognitive content (continued)</u> Let us now develop what we said earlier about my situation in W2. This will clarify and substantiate the points just made.

Remember that, in W2, I'm looking at somebody who looks just like Gary because it *really is* Gary, and I also correctly believe that thing to be Gary. I have pre-existing knowledge to the effect that Max, Gary's identical twin, is in another country and *also* to the effect that, apart from Gary himself, no person and no thing other than Gary looks just like him. Thanks to this knowledge, along with the information encoded in my current perception, I know that Max is the person who you're seeing.

Let VG be that perception. Further, let's continue to suppose that, from an experiential standpoint, VG is completely indistinguishable from the perception that prompted me to utter U1 and also from the perception that prompted me to utter U2. Finally, let's suppose that VG is the perception that

prompts me to utter U2.

Given these suppositions, it follows that from a purely experiential standpoint VG is completely indistinguishable from VM and also from VR. It also follows that the belief embodied in U2 is identical with the belief embodied in U1 and also in U3. U1-U3 all express the same belief. So CVR is the belief embodied in each of these three utterances.

Although I know that it's Gary I'm seeing, VG *itself* obviously doesn't tell me this. Nor does it tell me that, apart from Max, nothing in the universe looks just like Gary. These things were learned independently of VG. My knowledge of them is *background* knowledge.

Thus, there is no individual x such that x is identical with Larry and such that *by itself* VG tells me that it's x specifically that I'm seeing. I do indeed *know* who I'm seeing. But it's VG *plus* my background knowledge that gives this information, not VG by itself. Given only what VG is telling me, I could be seeing Gary or Max or some third thing, and the correctness of what I believe strictly on the basis of VG doesn't depend on who (or what) it is that I'm seeing.

3.0 <u>The gappiness of Burge-thoughts</u> The wrongness of Burge's position becomes painfully clear when we register it's most immediate consequences. Of those consequences, here's the one I find most striking. Given two people who are psychologically identical, one of them may be having all sorts of thoughts while the other is a zombie – a mental blank who is no more having thoughts than a tomato.[332]

Remember that, in W1, there is some individual x such that x=Max and such that U1(my utterance of 'that guy is wearing a plaid shirt') is true exactly if:

(MP[333]) x is wearing a plaid-shirt.

This is true. Notice that it's a claim about language, about literal meaning – about what is literally meant by an *utterance.* It isn't, at least not directly, a claim about the content of my visual perceptions or my thoughts.

But Burge holds that, indeed, positions exactly similar to the one just stated about that utterance hold with respect to my perceptions and thoughts. Burge's holds that, in W1, there is some individual x such that x=Max and such that the content of VM (the visual perception I'm having) is true exactly if

(MP[334]) x is wearing a plaid-shirt

Burge *also* holds that, in W1, there is some individual x such that x=Max and such that the content of what I *believe* on the basis of my visual perception is true exactly if MP.

This last claim has a very interesting corollary, as a short a extension of our comments about worlds W1-W3 makes clear. Let W4 be a world that, with one qualification, is like each of W1-W3 The qualification is that, at the time in question, Sally and I are hallucinating. In other words, I'm not really seeing a man who is wearing a plaid shirt; there's only empty space in the place where, so I believe, I 'see' there to be such a person, and Sally isn't having any such perception either. But I *am* having a hallucination that is phenomenologically (experientially) just like VM (and, therefore, VR and VG). We'll refer to that hallucination as 'VH'(short for 'visual hallucination'). And, prompted by that hallucination, I say:

(U4) ‘That entity is wearing a plaid shirt.’

If U4 were correct, it would be because, in the place in question (the one at which I’m pointing) there was some individual w, such that w was wearing a plaid-shirt; and the proposition that U4, if it were correct, would encode would be:

(bw) w is wearing a plaid shirt.

But U4 is not correct. It’s not true or false. It’s abortive. If I point to a non-existent gnome on my bed, and I say

(G1) ‘that gnome weights 8 lbs’,

my utterance *presupposes* that here is such a gnome, and since that presupposition is false, m statement is neither true nor false; it’s abortive. If I pointed to the bed and said

(G2) ‘there is a gnome on my bed, and that gnome weights 8 lbs’,

that statement would be false. (G1 and G2 are supposed to be utterances of sentences, not sentences per se – sentence-tokens, not sentence-types.) [335] But G2 is very different from G1 that’s a very different statement from the first one. U4 is abortive for the same reason. It is neither true nor false. And it therefore encodes *no* proposition. Anything that is true or false is *ipso facto* a proposition, and any proposition is *ipso facto* true or false. So U4 fails to encode a proposition.

Burge rightly believes that U4 fails to encode a proposition. But he

believes – partly on the basis, it seems – that *VH* has no content. In other words, he believes that VH isn't giving me *any* message – or, at least, that it's not giving me any *complete* message. The message it's giving me, thinks Burge, has a 'gap' in it. That message, Burge thinks, is like – nay, it *is* – the message encoded in U4. And *that* message is not complete (as Burge rightly believes), as it has a gap in it, thus having the form:

(b_) __is wearing a plaid-shirt.

The underlining (the ___) is meant to denote a gap or blank of some kind.

Burge also thinks that the though I'm having on the basis of VH has a blank in it – that it's like b__In Burge's view that thought (or non-thought, or whatever) is *identical* with b__

So, if we accept Burge's theory, I'm not really thinking *anything* in U4 – at least not to the extent that what am (or would be) thinking depends on my visual experience of the place in question.

Burge's position is preposterous. It's obvious that VH gives me complete message. It tells me that there is an individual in a certain place wearing a plaid shirt (etc.). That message is *false*, of course. But a precondition for being a false message is being a complete message.

___*is wearing a plaid shirt isn't true or false.* The expression '__is wearing a plaid shirt' (or just 'is wearing a plaid shirt') isn't true or false. It's just a predicate. It's no more true or false than 'north of' or 'lemon grass.' And the meaning of that predicate (which we may represent as: ___*is wearing a plaid shirt isn't true or false*) isn't true or false. But VH is true or false. It's false. It tells me that there's a plaid-shirt wearing man in a cetain place. There's no such man in that place. So VH is wrong. It isn't like G1. It's a complete, wrong message. And the thought I have on the basis of it is

wrong. That thought isn't (b_). It's

(CVR) There is some entity x such that x has such and such characteristics, such that x is standing is in a certain place and such that x is wearing a plaid shirt.

4.0 A corollary of Burge's externalism: the impossibility of knowing one's conscious thoughts any better than one knows facts about the spatially and temporally remote past Let's continue to think about my situation in W4. In that world, I *think* that I'm seeing somebody – and I'm wrong. But there's something I'm not wrong about. In W4, if, at the time I was having VH, I were asked 'are you having a thought and, if so, what it is?', I'd say without any hesitation: 'yes, I'm having a thought; I'm thinking that the fellow standing over there is wearing a plaid shirt.' Surely my answer to the second question is right. I know what I'm thinking. What I'm thinking is wrong, since there's no man there. But I'm right to think that I'm thinking it.

But Burge's position entails the opposite. In the context in question, I'm not thinking *anything*. There's a blank where, had I been in W1-W3, I *would* have been having a thought. A consequence is that, when you ask me 'are you having a thought and, if so, what it is?', I'd be *wrong* to say: 'yes, I'm having a thought; I'm thinking that the fellow standing over there is wearing a plaid shirt.' Why would I be wrong? Two reasons. First, I wouldn't be having a thought at all. I'd be having a would-be thought, a pseudo-thought, but not a thought. Second, since I wouldn't be having *any* thought, my belief as to what specifically I was thinking would also be wrong.

But that's obviously false.

4.1 <u>The causal conception of conception</u> To appreciate the depth and severity of this problem, we have to make a few points about Burge explicit.

According to Burge, in W1, the reason I'm thinking about Max, as opposed to Gary or someone else, is that I'm on the receiving end of a certain causal process that begins with Max. My current thought about Max (that he's wearing a plaid shirt, or whatnot) resulted from a perception of him. That perception obviously resulted from light bouncing off of him. Thus, light bouncing off of him caused me to have certain experiences, that at least part of the reason why those experiences are perceptions of Max, and that also being at least part of the reasoning my the thoughts in which those perceptions eventuated were thoughts about Max, and not Gary (or whoever). So my current thought – the one embodied in U1 – concerns Max, and no one else, because it 'tracks back' to Max, meaning that they bear a causal relation of the kind just described.

In Burge's view, conception is not to be understood in causal, not in descriptive, terms. Why does he thinks this? Remember Putnam's story about Smith and Twin-Smith. Smith's mental contents are *descriptively* just like Twin-Smith's. s identical. So far as he has a mentally encoded description that picks out H2O, it's 'liquid that we bathe in, drink, is found in lakes [etc.']', and so far as Twin-Smith has as a mentally encoded description that picks out XYZ, it coincides with Smith's. So given that, when Smith says 'water covers most of my planet's surface', he's making a statement about *H2O*, not XYZ, and that Twin-Smith says the same thing, he's making a statement about XYZ, not H2O, it can't be because Smith's thoughts are descriptively different from Twin-Smith's. The only relevant difference between the two is that Smith is, whereas Twin-Smith is not, on the receiving end of a causal chain beginning with H2O, whereas, Twin-Smith is, whereas Smith is not, on the receiving end of a causal chain beginning with XYZ, not

H2O. Of course, Burge believes the same thing *mutatis mutandis* to be true of Oscar and Twin-Oscar.

The essence of conception isn't description; it's the causal back-tracking relation just described. Smith is thinking about H2O because his thoughts track back to H2O, not XYZ. Supposing that you're right now thinking about Descartes, not Locke, that's because you're thoughts track back to Descartes, as opposed to Locke. I right now, am thinking about Descartes, and not Locke, because my thoughts

A story makes it clear how untenable this is. Right now it's 5:38 p.m., April 11, 2009). Let t our abbreviation for that time. Right now, at t, I'm thinking about Descartes. More specifically, I'm thinking:

(D*) Descartes was a fine philosopher.

I must emphasize that I'm doing a very a thorough job of reflecting on what's going on in my head right now; and, on that basis, I can say with complete confidence that D* is among the things that I'm thinking. Next year, historical research makes it clear that Descartes never existed – that there was no Descartes. It was all a hoax, we find out; the documents – all forgeries. According to Burge, I'm thinking about Descartes only in so far as my thoughts track back to him in the relevant way. They can't do that if he didn't exist. So when it's discovered that Descartes never did exist, it turns out, not only that, at t, I wasn't thinking D*. But I wasn't, if Burge is right.

If Burge is right, D* doesn't even exist to be thought. Supposing (for a moment) that Descartes *did* exist, there is some x such that x=Descartes, and such that D* is correct if, and only if, x is a fine philosopher and – Burge concludes form this -- such that, for that very reason, if x doesn't even exist, D* doesn't exist either. Burge thinks that Descartes is an actual *constituent* of

the proposition – of the (supposing for the duration of this sentence that Descartes did exist and was in fact a fine philosopher) *truth* – that he was a fine philosopher. (I myself don't think this. I discuss this in Chapter__, where I identify the fallacies concerning the nature of propositions that underlie this view of Burge's.)

So, supposing (as we were doing a moment ago) that Descartes turns out never to have existed, despite years of thought to the contrary, it will turn out, in Burge's view, that D* doesn't even exist to be thought. So I can't even think that I'm thinking it – *it* doesn't exist, after all.

And that's a pretty strange view.

But it gets worse. Consider the various mental events and brain events (and mental and neural conditions etc.) involved in (what I thought was) my failed to attempt, at time t, to think about a certain French philosopher. If Burge is right, it turns out, not only that I wasn't thinking D*, but that the aforementioned neuro-mental states of affairs didn't constitute *any* thought. I was shooting a cognitive blank.

4.1.2 Burge's response[336] Burge became of aware of this problem quite some time ago. He denies that his position is inconsistent with the presumption that people know what they are consciously thinking. Let's start with what Burge admits. Burge admits three things, which we'll refer to 'B1', 'B2', and 'B3':

(B1) We can't know for sure that Descartes existed. No matter what mental states I have, it's a possibility that he didn't. The documents on which my belief is based could have been forged (etc.).

Thus, even though (so I think, and so let us assume), I *do* know that Descartes existed (at least in so far as I know anything about the what

happened in France centuries ago), I don't know *for sure* that he did. Moreover, given only what my own mental states are, it's a theoretical possibility, that he never did exist.

(B2) For argument's sake, suppose that, in reality, Descartes really did exist. Let's also suppose that I'm now thinking:

(D*) Descartes was a fine philosopher.

A consequence of Burge's position that there existed some individual x, namely Descartes, such that D* is correct if, and only if,

(D1*) x was a fine philosopher

and such that the mental events (or structures or whatever) involved in my having thought are correct if, and only if, D1* is correct.

(B3) B2 entails that *if* I were to be in the very same mental state I was in when I was thinking D*, *but* it happened to be the case that Descartes didn't exist, then I would *not* be thinking D*. In other words D2 entails that the very mental events and brain-events that *in fact* constituted my thinking D1* would fail to constitute *any* thought. By virtue of having those mental and neural events, I'd be shooting a blank, not having a thought – I'd no more be having a *bona fide* thought than a tomato.

B1-B3 are all correct. B1 is correct, since I can't know for sure that my mental states correctly represent the external world or *a fortiori* that

inferences involved in my believing in Descartes existence are correct – inference that are made on the basis of possibly faulty perceptions of possibly faulty records concerning 17th century France. B2 is correct, since Burge's position *does* entail Descartes must have existed if anyone is to think that he was a fine philosopher. And B3 is correct since Burge's position *does* entail that the very mental events involved in my thinking D1* would fail to constitute my having any thought at all if, supposing that everything else were the same, Descartes had never existed.

B1-B3 entail that:

(B4) If I'm now thinking that

(D*) Descartes was a fine philosopher,

I can't know it with any more certainty than I can know that Descartes existed.

But that's wrong. I obviously know what I'm consciously thinking with greater certainty than I know about medieval France.

I want to make it as clear as possible why B1-B3 entail B4. Given B2 , I can think D* only to the extent that Descartes really did exist. Burge admits this. (He has to. B2 is Burge's whole position.)

Given B1, I can know that he existed only to the extent that various beliefs of mine concerning the external world to be accurate – beliefs that presuppose the accuracy of many perceptions of mine and also of many historical documents (concerning France in the 17th century). Burge admits this as well.

B3 entails that three is no experiential difference – no difference that is

even theoretically detectable – between how I am when I'm *actually* thinking D*1 and how I am in a world where Descartes never existed but is otherwise like the world where I'm actually thinking D*1.

B3 entails that, if I am to know that I'm thinking D*1, or anything else about Descartes, I'd jolly well be know that he existed; and I'd therefore jolly well better know that various perceptions of mine were accurate and, in addition, that various inferences of mine concerning various records of mine were also accurate. Which means that a *pre*condition for m thinking about Descartes is my knowing that Descartes existed Which, in its turn, means that, supposing that I'm consciously thinking about Descartes right now, I can't possibly know that fact with very much certain at all. I can know it only with the certainty with which I know that Descartes existed.

But that's false.

Since knowing that Descartes existed obviously involves a lot of other things about the external world – it involves knowing that I was in fact reading certain books at certain terms, that certain records in fact exist and were made under certain circumstances, etc. – my knowing a huge amount about the external world is a pre-requisite to knowing that I'm thinking this or that about Descartes.

But that's false.

This reasoning is easily generalized to show that, given anything x that is in the external world, my knowledge that I am consciously thinking about x is no more certain than my knowledge that x existed and, therefore, is no more certain than my knowledge of the various things on which that belief is based (which, where almost anything concerning the external world is concerned, is vast).

But that's false. I know what I'm consciously thinking. Burge's position entails otherwise. Burge's position is wrong.

Burge is aware of this problem. He became aware of it around six years after he wrote 'Individualism and the Mental.' He and various followers of his have spent the last twenty-three years attempting to show that, even though B1-B3 are all correct, Burge's position is completely compatible with the presumption that we know our conscious mental states better than we know about Descartes existence.

To those[337] who would join Burge in this heroic quest, I must say, quoting Sponge bob square pants and : 'Good luck with that!'

5.0 <u>The consequences of Burge's big blunder: the need to deny independently known truths about causation</u> But there are problems with Burge's position far worse than any thus far mentioned. To see what they are, let's revisit the situation involving Smith and Twin-Smith – the one where, according to Burge, one of them, but not the other, believes that he has arthritis.

By supposition, Smith and Twin-Smith are psychologically and physiologically exactly the same. Given this, it follows that the causal properties of Smith's thoughts coincide with those of Twin-Smith's thoughts. This means that the suppose difference in content doesn't make a difference. At all. It is completely and totally inert. Although Smith's *belief* in the proposition expressed by A3 (or whatever) obviously has effects (on what Smith does, etc.), and although the same is obviously true of Twin-Smith's corresponding belief, the alleged *difference* between these beliefs is absolutely without any consequences. It has no consequences for anybody's overt behavior or for anybody's internal, psychological condition.

5.1 <u>Fodor's insight</u> Jerry Fodor (1987) was one of the first people to note this consequence of Burge's view. Fodor's argument is (characteristically for

Fodor) a brisk and incisive one. What is it, Fodor asks, for two numerically distinct things, x and y, to be causally different, i.e. to differ in respect of their causal properties? It is for them to be such that, were they to switch places with one another but all other factors were left unchanged, the one would behave in the way in which, in actuality, the other behaves.

Here's the idea. Suppose that x and y are two billiard balls that are atom-for-atom duplicates of each other. At time t, x collides with a green billiard ball and causes that billiard ball to move. At the same time, y collides with a red billiard ball and causes that billiard ball to move.

It's clear that the *effect* of the one collision are different from those of the other. In the one case, the effect is to make a green ball move; in the other, it is to make a red ball move. But this doesn't mean that the two billiard balls differ in their causal properties. It means that a given causal property has different effects in different contexts. Suppose that I'm moderately strong. In some contexts, that will have the effect that I easily lift the barbell I'm trying to bench-press. (Such contexts include those where the barbell weighs under 100 lbs.) In other contexts, that will have the effect that I can barely make the barbell budge. (Such contexts include those where the barbell weighs over 350 lbs.).

But, assuming that my level of strength hasn't changed in the intervening time-period, the difference between a situation where I'm trying to lift 50 lbs and one where I'm trying to lift 350 lbs isn't that my causal properties have changed; it's that one and the same set of causal properties are sufficient for certain jobs but insufficient for others. (You can't make a rocket out of plywood. But you can make a bird-house out of plywood. That doesn't mean that the causal properties of ply-wood change, depending on whether they're composing a rocket or a bird-house. It means that those causal properties, though sufficient for certain purposes (e.g. bird-house

building) are insufficient for others (e.g. rocket building).

In his landmark work, *A system of Logic*, John Stuart Mill made it clear that the just-described criterion of causal identity is (i) presupposed by all experimentation and (ii) is, setting aside the empirical fact that scientists rely on it, correct as a matter of logic.

This is not to detract from Fodor's achievement. Fodor stated that criterion more sharply and clearly than any author that I know of who preceded him. Plus, Fodor deserves credit for seeing the bearing of this principle on Burge's views. Fodor deserves a lot of credit for a lot of positive developments in recent analytic philosophy, as we'll soon see.

5.1.2 <u>Why thoughts must have causal powers</u> Causality is indistinguishable from spatiotemporal existence. To spatiotemporal is to occupy some region of space-time. It makes no sense to say that x occupies a certain region of space time but has no causal properties, since occupancy is itself a causal notion. My occupying the chair I'm in is indistinguishable from my altering the conditions that other things, e.g. you, must satisfy if they are to occupy the chair. If I weren't in the chair, then (other things being equal), you wouldn't have to displace me to occupy the chair. Given that I am occupying the chair, you would have to displace me to occupy it yourself. And that, though certainly possible, would be a hassle for you.

Given that Smith and Twin-Smith are absolutely the same in respect of their causal properties, it follows that the alleged differences between the one person's mental content and the others is causally impotent. A moment ago we saw that it makes no sense to posit the existence of anything spatiotemporal that has no causal powers at all. The just-cited alleged difference has no causal powers. Therefore it makes no sense to supposed that it exists. Therefore it doesn't exist. Therefore Burge is wrong.

5.2 <u>Burge's response to Fodor</u> Burge's response to Fodor is rather curious. Burge *denies* that billiard balls x and y are causally identical. Burge says that the causal properties of an object are to be understood in terms of what it *actually* does, and not at all, contrary to what Fodor says, in terms of what it *would* do in different circumstances. On Monday, let us suppose, I run a six minute mile without wearing a twenty pound backpack. On Tuesday, a run a seven minute mile wearing such a backpack. On both days, I run on the same path. The weather is the same. I'm no worse and no better rested or nourished on the one than on the other. Etc. According to Burge, my level of fitness has declined, the reason being that I ran the mile more slowly on Tuesday than on Monday. It's irrelevant, says Burge, that, on Tuesday, I *would* have run the mile as quickly as I did on Monday *if* I hadn't been wearing the backpack. Why is it irrelevant? Because, says Burge, the causal properties of physical objects are to be understood in terms of the peculiarities of the environments in which they happen to be embedded.

Let's look at the example that Burge himself uses. Let A and B be two celestial bodies that are atom-for-atom duplicates of each other. A is revolving around a star. So A is a planet. B is not revolving around any star. So B is not a planet. Burge says (rightly) that the concept of a planet is an explanatorily important one. We explain at least some things in terms of the fact that some things are planets. From this, Burge concludes that A and B have different causal properties. It is not, says Burge, that A and B merely differ in their behavior. It is that, by virtue of being a *planet* – albeit one that is exactly like B down to the last atom – A has causal virtues that B lacks.

Is Burge's argument a good one? No. Suppose that Smith and Jones are atom-for-atom duplicates of each other, but that Smith has a nephew whereas Jones does not. Obviously there are, or at least could be, behavioral

differences between Smith and Jones that that are to be explained in terms of the fact that the one, but not the other, is an uncle. Why did Smith spend an hour shopping for toys?' Because he's six-year old Timmy's uncle, and Timmy loves toys, as Smith knows. Why didn't Jones spend an hour shopping for toys? Is it because Jones is mean and stingy, whereas Smith is kind and generous? No. It's because Jones, unlike Smith, isn't an uncle and thus does not (at least not for that reason) have go toy-shopping. Burge's position concerning planets, if transposed to this other context, is that Jones is in fact a very different person from Smith. Smith is a nice guy. (Let's assume for argument's sake that his toy-shopping is an expression of genuine love for his nephew.) But Jones, not having gone to the trouble of buying toys for his non-existent nephew, is not such a nice guy, even though Jones is just as likely as Smith to buy a present (or otherwise behave amicably towards) the people in his life who, unlike his nephew, actually exist.

Let's get back to A and B. Because A is embedded in a certain way in a certain solar system, it's going to do things that B won't do. It'll move in ways in that B will not. It'll draw the astronauts who land on A towards itself, whereas B will not do that, since (we may supposed) no astronauts bothered to go to B.

This is all very true. But does it entail that A has different causal properties from B? No. In describing A as a 'planet', one is saying that it is embedded in a certain way in a certain environment. In describing B as a non-planet, one is one is saying that it is not embedded in that way in that sort of environment. Burge is quite right what things do depends on the circumstances. If I'm underwater I don't breathe; if I not, I do. Since A is part of a solar system, it does certain things; since B isn't, it doesn't. But this doesn't show that A and B differ in respect of their causal properties from each other. It shows that:

(i) A's being embedded in a certain environment in a certain way

has different effects from

(ii) B's being embedded in a totally different way in a completely different sort of environment.

(i) is not identical with A. (ii) is not identical with B. (I am sitting in a chair right now. But I'm not identical with *the state of affairs of JMK's sitting in a certain chair at 1:22 p.m. on April 9, 2009*.) A's being planet is identical with (i)'s being the case. (To be a planet *is* to be a thing of certain dimensions (etc.) that is embedded in a certain way in a certain sort of environment.) Since A's being a planet is no more identical with A than I am identical with the fact that, at this instant, I'm sitting in a certain chair, and since the same thing *mutatis mutandis* holds of B, it would be deeply illogical to accept Burge's judgment that A and B differ in their causal properties. In order for that judgment to be correct, A *would* have to be identical with the fact that it was embedded in a certain way in a certain sort of environment, and by parallel reasoning you *would* be identical with the fact that, right now, you are holding a certain book in your hands. But so far as it isn't obviously false, this position borders on meaninglessness.

Despite all these problems with content-externalism, and despite some comparably gruesome ones that I haven't even mentioned[338], content-externalism still has some loyal followers, who are indefatigable in their attempts to make content-externalism work. To these tenacious individuals, I can only say, quoting the great Marion Barry: 'Let it go!'

The Computational Theory of Mind

These days, a doctrine known as the "Computational Theory of Mind" (CTM) is very popular, both among scholars (especially philosophers and computer scientists) and among the laity. CTM is a form of materialism, for reasons that will become clear.

One would think that this doctrine would be informed by the same intellectual progressiveness that led to the creation and subsequent development of computer science. And, because of its liberal and confident use of words such as "compute" and "algorithm," CTM has an impressive sheen of rigor.

But we will find that, ironically, CTM involves a regression to archaic, animistic forms of thought---that, more specifically, it represents the recrudescence of the primitive, and spurious, thinking that the grumbling of thunder represents the wrath of the gods.

Speaking more philosophically and less psychoanalytically, we'll find that CTM involves some basic ratiocinative failings. For example, CTM embodies an ignorance of the fact that it is events that have effects, not objects. These events often involve objects. But objects *per se* cause nothing. *The rock* didn't break the window; what did so was the *rock's striking the window* (with a certain degree of force). We'll also find that, if CTM were true, we couldn't know it to be true and couldn't have any legitimate reason to advocate it.

1.1 What is CTM?

According to CTM, to think is to compute. Thoughts are computations. A computation is a formal inference. So CTM is the view that any instance of

thought is identical with the making of a formal inference.

But what is a "formal" inference? To qualify as "formal," a case of reasoning must be identical with a reaction, on the part of some creature or other, to symbolic-inputs. But not all such reactions qualify as cases of formal reasoning: only those reactions so qualify that aren't to any degree driven by the knowledge of the *meanings* of the symbol-tokens in question and are driven entirely by a sensitivity to their *forms.* So CTM is the contention that thought consists of responses to symbol-tokens that are driven entirely by the formal properties of those tokens and therefore embody no knowledge of what those symbol-tokens mean. In Section 1.2, we'll give some examples of such responses, thereby eliminating any obscurity from the statement just made.

Why advocate CTM? Two reasons. First, CTM, if correct, solves the mind-body problem. When *you* add two numbers, you are *ipso facto* thinking. Computers presumably add numbers. (Computers presumably compute.) Therefore, computers, when adding numbers, are *ipso facto* thinking. We know how a computer's software is related to its hardware. So CTM, if correct, solves the mind-body problem. (In other words, CTM, if correct, makes it clear how brains are responsible for the mental activity that they mediate.)

Second, CTM, if correct, solves the "tracking problem" (or so CTM-advocates say and so we'll momentarily assume). Brain-activity, though clearly 100% physics-driven, is also logic-driven (or, in any case, logic-*consistent*), at least to the extent that it mediates rational thought. So, to the extent that it's constitutive of rational thought, brain-activity, though governed by *de*scriptive laws, embodies sensitivity to *pre*scriptive laws: where such activity is concerned, the descriptive "tracks" the prescriptive. Nothing that is obeying the laws of physics is, in virtue of that fact, doing

anything that can be *evaluated*--- that can be described as “correct” or “valid” or “false.” (In moving hither as opposed to thither, the billiard ball isn’t doing anything right or wrong, valid or invalid.) And, it seems to follow that nothing that can be so described could be constituted by brain-activity. But brain-activity *does* constitute activity that can be so described.

CTM (supposedly) solves this problem. It’s possible to create an environment in which there occur symbol-tokens whose logical properties are hewed to their physical properties in such a way that, within that environment, symbol-tokens occur in sequences that are consistent with the requirements of logic. In fact, not only is it possible to create such environments: we routinely *do* create them—and employ them. We call them “computers.” So it would seem that CTM, if correct, explains how activity that falls within the scope of physics can, at the same time, embody a sensitivity to the demands of logic and, more generally, of normativity.

1.2 Mentalese: the “language of thought”

According to CTM-advocates, those symbol-tokens mediating our thoughts are held to belong to a hypothetical language. This hypothetical language is generally referred to as “Mentalese.”

According to CTM, one doesn’t have to *learn* Mentalese. Brains are computers. Computers don’t function until they’ve been programmed. Therefore, our brains don’t function until they’ve been programmed. Mentalese is that program. Thus, Mentalese must be programmed into a brain *before* that brain can mediate cognitive activity, and therefore *before* that brain’s owner can learn a language or, indeed, do anything that involves having any awareness of anything. Mentalese is to a functioning brain what a computer’s operating system is to that computer. Thus, just as your IBM

came “bundled” with DOS, so your brain came “bundled” with Mentalese.

In light of what we’ve said, it might seem reasonable to say that, according to CTM, one innately *knows* Mentalese. But if CTM is right, one doesn’t know Mentalese at all, let alone innately. Knowing a language involves being able to correlate linguistic with non-linguistic entities. If I couldn’t correlate snow (the powdery substance) with “snow” (the word), I couldn’t possibly know what “snow” meant. In order to correlate “snow” with snow, I must have some way of being aware of snow that doesn’t involve knowing what “snow” means. Knowing Mentalese would involve being able to correlate the expressions belonging to it with non-linguistic entities. According to CTM, thoughts are identical with Mentalese expression-tokens, and such expressions must therefore be operational if one is to make correlations or, therefore, to know a language.

So if CTM is right, a *pre*condition for one's being aware of anything (or making correlations of any kind or, knowing a language) is that one's brain be stocked with already operational Mentalese symbol-tokens. So it isn’t by virtue of one’s *knowing* Mentalese that expression-tokens belonging to it mediate one’s thoughts or otherwise do what CTM-advocates allege them to do.

To sum up: Nobody knows Mentalese; nobody understands the expressions belonging to it; and it isn’t to any degree being understood that such expression-tokens facilitate thought.

In this respect, Mentalese is radically different from English and every other non-hypothetical language. Expression-tokens belonging to English, Urdu, etc. do not facilitate thought *except* by being understood. By contrast, the symbol-tokens of Mentalese, supposing such tokens to exist, need not be understood to facilitate thought. Indeed, it is inherent in what Mentalese expression-tokens are that they be able to discharge *all* of their duties *without*

being understood to any degree by anyone or anything.

Because Mentalese, were it to exist, would differ in this respect from English, Spanish, and other *bona fide* languages, it follows (for reasons to be discussed in Section 4.0) that the very concept of a "language of thought" is not a coherent one. Given that CTM assumes there to be a language of thought, and given that (for reasons to be disclosed in Section 4.0) no doctrine that doesn't assume this bears any resemblance to CTM, it follows that CTM is incoherent.

1.3 The concept of a formal inference: Some illustrations

We've seen, *if* correct, CTM explains much. But *is* it correct? We can't answer that unless we make it clear what, in this context, the word "compute" means. And we can't make *that* clear unless we make it clear what, in this context, it means to say of symbolic activity that it is "form-driven."

The following three stories will help answer these questions.

Story #1 (*not* an example of computation) Smith, who speaks English, is told: "Jones has two yachts." On the basis of this information, along with his knowledge of the English language, Smith concludes that the following sentence is correct: "the number *n* such that Jones has *n* yachts is such that *n* has no rational square root."

Commentary on Story #1: Smith's reasoning is cogent. But it is not in any sense "formal." Smith's thought process was a driven by an understanding of *inter alia* the *meanings* of the words he was hearing.

Story #2 (an example of *a* kind of computation) *W* is a wood-chipper. But

W is no ordinary wood-chipper. Suppose that *W* is fed, in the following order, a wooden "2," a wooden "+," a wooden "9," and, finally, a wooden "=." Suppose, in other words, that *W* is fed an instance of the sentence-fragment "2 + 9 =." In that case, driven *solely* by the morphological properties of the symbols that it was fed, *W* cranks out a wooden "11." And, *W* does the same thing *mutatis mutandis* for any other two numbers. So given any whole numbers *M* and *N*, if *W* is fed the sentence-fragment "*M* + *N* =," then, driven solely the morphologies of that fragment, *W* produces a wooden sign denoting the sum of those two numbers.

Commentary: There would seem to be *a* sense in which *W* can compute sums. It is typically *this* sense of the word "compute" that advocates of CTM have in mind when they say that to think is to compute. Why only "typically"? Why not "always"? Because sometimes what they have in mind is the *prima facie* quite different sense of the word "compute" that is illustrated in the following section.

Story #3 (an example of a different kind of computation) Smith has no idea what "1,245," "241," "1,486," etc. mean. But Smith has been told by a source whose testimony he rightly trusts that, when given ink-marks having such and such shapes, a correct sentence results when another ink-mark, having thus and so shape, is placed in a certain position relative to the original ink-marks. In particular, Smith knows that, given their respective shapes, the ink-marks "1,486" and "1,245 + 241 =" form a correct sentence when (some isomorph of) the former is placed to the right of (some isomorph of) the latter. It must be emphasized that Smith has no idea what any of these ink-marks mean. Smith knows nothing of relevance to this situation *other* than the rules as to how certain shapes ought to be manipulated or

complemented. Those rules *in fact* describe how to do sums involving multi-digit numbers. But Smith doesn't know that. He knows only that, when given ink-marks having such and such a shape, another ink-mark must be placed beneath them if a true statement (as to whose identity he is quite ignorant) is to result.

Commentary: What Smith is doing seems to be different from what *W* is doing. Smith's applying the just-mentioned rule involves his having various sense-perceptions (e.g., he sees various ink-deposits) and various beliefs (e.g., he believes that those ink-marks are symbols of some kind or other, and he believes that a correct statement, albeit one of whose meaning he is ignorant, will result if, underneath them, he produces an ink-mark having a certain shape); and his having such beliefs involves his making quite delicate judgments (e.g., he must judge that the ink-deposits in question were produced by sentient beings with certain intentions---he would *not* see those ink-deposits as tokening the sentence-fragment "1,245 + 241 =" were he to learn that their existence was the unintended result of some explosion involving an ink-filled urn).

Nothing comparable holds of *W*. It isn't because of any sense-perceptions or beliefs or intentions on its part that, having been fed a wooden token of the sentence-fragment "2 + 9 =" *W* spits out a wooden "11." So what Smith is doing in writing a "1,486" seems *not* to be comparable to what *W* is doing in spitting out a wooden "11."

And, while there is certainly *a* sense of the word "compute" in which Smith is computing a sum, this cannot possibly be the sense in which, according to CTM, to think is to compute. A *pre*condition for Smith's performing the just-described computation is his *already* being able to upload and process information. So CTM would be guilty of vicious circularity were

it given by the thesis that, in *this* sense of the word "compute" to think is to compute. So, supposing CTM coherent, the sense in which to think is to compute is the sense in which *W* computes, and the sense in which *Smith* computes is irrelevant.

We will now see that CTM is indefensible and that no doctrine at all similar to it can possibly be correct. There are five reasons for this. These reasons will given in Sections 2.0-6.0.

2.0 CTM's failure to distinguish between shape and syntactic form

A story will help expose one reason why CTM is false. You shout the following at me: "if your nose starts to bleed in the next ten seconds, I'll give you a million bucks." Let *T* be the sentence-token in question. As a result of *T*'s being uttered, I *do* start to bleed out of my noise, and you duly give me $1,000,000. But it wasn't because of what *T* meant that I had a nose-bleed; nor was it because of *T*'s syntax. It was *T*'s being loud that caused my nose-bleed.

My reaction here wasn't *symbolic.* Even though *T* is a symbol (or, to be precise, a symbol-token), *T*'s being a symbol did nothing. For exactly similar reasons, *W*'s spitting out an "11" isn't symbolic, even though its doing so was precipitated by events involving symbol-tokens.

2.1 This line of reasoning generalized

We've just seen that CTM is a non-starter if a symbol-token's "form" is identified with its shape. But sometimes, when referring to a symbol-token's "form," advocates of CTM are (so they say) referring to its *syntactic* form. Given this use of the word "form," CTM becomes the contention that

thoughts are syntax-driven operations on symbols.

This contention is easily refuted. Syntax is semantic structure. “Smith loves Mary” and “Sally hates Greg” have the same syntax because the semantics of the one has the same *structure* as the semantics of the other.

To be sure, those sentences have very different meanings. But they’re semantically isomorphic. Each constituent of the one sentence corresponds to exactly one constituent of the other (“Smith,” “loves,” and Mary” correspond to “Sally,” “hates,” and “Greg,” respectively); and the relationship borne by the semantics of any given constituent of the one sentence to the semantics of that sentence as a whole coincides with the relationship borne by the semantics of that constituent’s counterpart in the other sentence to the semantics of *that* sentence as a whole. So the *relation* borne by the semantics of “Smith” with respect to the proposition that Smith loves Mary coincides with the relation borne by “Sally” to the proposition that Sally hates Greg.

Given that syntax is semantic structure, syntax-driven operations are *ipso facto* meaning-driven. Being the contention that “formal” operations are *not* meaning-driven, CTM is therefore completely incoherent if taken to be the contention that “formal” operations are syntax-driven operations.

At the same time, CTM is incoherent, as we saw, if “formal” operations are taken to be shape-driven operations. There is no third way to interpret the word “formal.” So CTM is incoherent.

2.2 Shape vs. syntax (revisited): The equivocation that lies at the heart of CTM

Here is the thinking embodied in CTM:

Form-driven responses to symbolic inputs are *ipso facto* cases

> of thought. *W* has form-driven responses to symbolic inputs. Therefore, *W* is thinking (it is adding numbers).

But when this argument is stated in a clear and non-elliptical manner, it becomes clear that it involves an equivocation and is invalid for that reason:

> Form-driven responses to symbolic inputs are *ipso facto* cases of thought. (In other words, *syntax-driven* responses to symbol-tokens are *ipso facto* cases of thought.) *W* has form-driven responses to symbolic inputs. (In other words, W's responses to its inputs are *morphology*-driven.) Therefore *W* is thinking.

The second argument is clearly spurious. Presumably because they are aware of this, advocates of CTM typically hold that, even though a symbol-token's shape isn't *identical* with its morphology, syntax can be "coded" into morphology---that syntax can be "morphologized" in much the way that geometry can be arithmetized. This is both false and irrelevant.

It's false because anything, no matter what its shape, can have any syntactic structure, and anything, no matter what its syntactic structure, can be tokened by anything, no matter what its shape. (In other words, given an arbitrary shape, and given an arbitrary syntactic structure, there is a possible language to which there belong expressions having that syntactic structure that are tokened by objects having that shape.) It's irrelevant because, even if syntax can be morphologized, it isn't, as we've already seen, *in virtue of* its having this as opposed to that shape that a symbol-token has this as opposed to that syntax. So *in so far as* one is responding to an object's shape, one's behavior is, to that extent, *not* the effect of, or otherwise to be understood in terms of, that object's shape.

To be sure, one knows what syntax an object has *on the basis of* its shape. A large part of the reason that I know the syntax of the following sentence-token---"John is tall and grass is green"---is that I know what shape that ink-deposit has. But sentence-tokens with precisely the same syntax can be tokened by hand-movements *or* by noises; and hand-movements and noises have no strictly morphological resemblance at all to any ink-deposit. (No ink-deposit that tokens "John is tall and grass is green" bears much resemblance to a stream of noises, or a series of hand-movements, that also tokened that same sentence, except relative to rules of projection that were contrived *ex post facto* to establish such a resemblance.)

Also, the relationship that holds between (a) my knowledge of that ink-deposit's shape and (b) my knowledge of its syntax is of a *causal*, and not of a *constitutive*, nature. I see that it has that shape. Given that knowledge, along with my background linguistic knowledge, I judge that it has a certain syntax.

To be sure, from an experiential (phenomenological) perspective, it may seem as though *in* knowing that it has that shape I *ipso facto* know that it has that syntax. But that's not the case. What we have is one bit of knowledge (knowledge of shape) *leading*, by way of (rapid) judgments, to some other bit of knowledge (knowledge of syntax).

Echoing what we just said, these two bits of knowledge may seem, at the level of consciousness, to come into existence simultaneously. But that appearance is misleading and, even if not misleading, irrelevant. It's misleading because any knowledge of an object's syntax *must* follow knowledge of its physical properties. To clarify these last two statements: If I judge that you're angry, it's on the basis of your facial expression, or some such. My judgment as to your emotional state may immediately follow my awareness of your expression. But given that the former is *based* on the latter and is also *caused* by it, the one awareness follows the other, however

quickly. Exactly similar points hold of the relationship between my knowledge of a given object's physical shape and my subsequent knowledge of its syntax: it isn't possible for me to have the latter knowledge without having the former; it *is* possible for me to have the former without having the latter; and my knowledge of the former is the sole foundation for, and is also what leads to, my knowledge of the latter.

Plus, it's irrelevant that, speaking from a strictly experiential (phenomenological) perspective, the occurrence of the one bit of knowledge *seems* to coincide with the occurrence of the other. It is the purpose, or at the very least the effect, of consciousness to condense the intricacies of the cognitions underlying it into a single, seemingly homogeneous "experience." If every articulation of pre-conscious cognition were faithfully represented in consciousness, if phenomenology faithfully duplicated every joint in the intricacies of the cognitions of which our conscious judgments are the outcomes, then consciousness would be so overburdened with information that it would cease to be able to perform the executive functions with which it is tasked. When seeing a rabid dog run towards me, my consciousness would present to me a richly articulated mosaic of judgments, whose consequences for my well-being would take much more time to figure out than was available to me. Fortunately, however, my consciousness presents me with a homogenous feeling that causes me, not a moment too soon, to bolt into action. Instead of giving me the psychological equivalent of a two-hundred page state-department report, consciousness gives me the equivalent of an electric cattle-prod in my back. And because the rich cognitive back-story is telescoped into that bolt of unpleasant and, aside from its distinctive affective nuances, homogeneous bolt of unpleasant phenomenology, I do the right thing (i.e., I run to the mini-van, lock the doors, roll up the windows, etc.) and, what is more, I do it in real time.

In general, it is precisely *because* consciousness represents as homogenous "feelings" what are in fact intricate networks of judgments that the contents of consciousness, unlike their cognitive underpinnings, are *actionable.* It is, I would conjecture, the very purpose, or at least one of the purposes, of consciousness to condense cognitive intricacies into affective simplicities. Therefore, it is *inter alia* its purpose to condense distinct, albeit adjacent, stages of intricate thought processes into a single, action-causing, survival-ensuring *feeling*.

In any case, if this is at all on the mark, then it's of no relevance that my recognition of the ink-deposit's shape *seems* to be temporally coincident with my recognition of its syntax. It is to be pointed out that, if this *isn't* on the mark, then much work in psychology crumbles.

3.0 A related point: CTM incompatible with the fact that distinct thoughts can be part of some one thought-process

If CTM is right, my thinking *2 is a prime number and 2 is less than three* cannot be responsible for my later thinking *some prime number is less than three.* According to CTM, if B_1 is the totality of brain-events constitutive of my thinking *2 is a prime number and 2 is less than three*, and B_2 is the totality of brain-events constitutive of my thinking *some prime number is less than three*, it cannot be B_1's *being* such a belief that brings about B_2. So it cannot be B_1's being such a belief that leads me to believe that some prime number is less than three.

So, if CTM is right, B_1's being such a belief has nothing to do with B_2's existence or *a fortiori* with B_2's being constitutive of a thought to the effect that some prime number is less than three. Thus, my believing that two is even and is less than three cannot, if CTM is right, have anything to do with

my later believing that some even prime is less than three. So, if CTM is right, T_1 has nothing to do with T_2, notwithstanding that the brain-events constitutive of T_1 brought about those constitutive of T_2.

So even though B_1 occurs right before B_2 and is causally responsible for it, and even though B_1 and B_2 are constitutive of, respectively, T_1 and T_2, nonetheless T_1 and T_2 cannot, if CTM is right, be installments in some one thought-*process.*

In general, CTM is incompatible with the fact that distinct thoughts can be part of some one thought-process. CTM is thus incompatible with the fact that thought-processes occur---with the fact that there are *sequences* of thoughts (setting aside such thought-sequences whose constituents have no thematic coherence, having at best the content-free coherence had by a series of domino-fallings).

4.0 CTM's dependence on the false assumption that syntax can mirror logical form

Context-sensitivity drives a wedge between syntax and logical form. If I say:

(1) "it is snowing *now* and it is not snowing *now*,"

what I'm saying might be true---it might stop snowing before I utter the second conjunct---even though it has the same syntax as:

(2) "*n* is even and *n* is not even."

Similarly, if I say:

(3) "that guy is identical with that guy,”

what I'm saying might be false, even though it has that same syntax as

(4) "*n* is identical with *n*."

If our thinking were syntax-driven, we couldn't recognize the differences in respect of logical form between (1) and (2) or between (3) and (4). We can recognize those differences. So our thinking is not syntax-driven. So CTM is wrong.

Here is a similar line of thought. If, on July 1, 2010, in Richmond, VA, I say:

(i) "right now, Jim is three thousand miles from here,"

the information encoded in my utterance is very different that encoded in an utterances of:

(ii) "On July 1, 2010, Jim is 3,000 miles from Richmond VA."

The information encoded in (i) doesn't follow from the information encoded in (ii), or *vice versa.* In general, when the context-sensitive expressions in a true statement are replaced with context-*in*sensitive expressions, the result is a statement that, although true, neither entails the first statement not follows from it.

For an expression to be context-sensitive is for its *referent* to vary with the

context. What an expression refers to is determined by its semantics, not its syntax. CTM says that thought is to no degree semantics-driven. If that were true, we couldn't grasp perspectival truths, such as that expressed by (i). Since we can grasp such truths, CTM is false.

5.0 CTM's reliance on the erroneous assumption that there are minimal units of thought

First a definition: a "morpheme" is an expression that cannot be decomposed into others. Bearing this in mind, let's suppose that, just as CTM says, we think in a language. In that case, our concepts are identical with expressions. Linguistic expressions consist of morphemes. (Any given complex expression; e.g., "the cat is on the mat," decomposes into expressions that cannot, in their turn, be further decomposed; (e.g., "cat," "mat"). So in contending that our thoughts are identical with (tokens of) Mentalese symbols, CTM is, in effect, assuming that we have simple concepts: concepts that cannot be decomposed in others. So CTM, if correct, entails that any given thought that one has can be decomposed into "simple" concepts. CTM thus has the radically implausible consequence that we have "simple" concepts. In other words, CTM has the consequence that some of our concepts are no more capable of being decomposed into other concepts than a morpheme (e.g., the word "cat") can be decomposed into other expressions. If this consequence of CTM is correct, it follows, as we'll now see, that some of our concepts would be such that one could have them without also having other concepts.

Let *T* be the token of the Mentalese word for "snow" that went off when, just a moment ago, I thought to myself that snow is white. We may assume

that *T* is a morpheme. (If, for some reason, you reject that assumption, just map what we say about *T* onto some Mentalese expression-token that, in your judgment, is a morpheme—and you must, if you grant that we think in a language, grant that there is some such token, since morpheme-less language is a surd---and everything that we say about *T* will hold *mutatis mutandis* of this other expression-token.) Not being decomposable into other expression-tokens, *T* presumably exists independently of such tokens. Given any *simple* expression—given any morpheme—belonging to a language *L*, you can remove that morpheme without removing other morphemes. Of course, in removing that morpheme, you'll remove all of the expressions of which it is a constituent. (And, if you choose a non-lexical item; e.g., "and" or "of," you may thus do a lot of damage to *L*.) But you won't, in removing that morpheme, be removing *other* morphemes.

Therefore, if, as CTM alleges, we think in a language ("Mentalese") , some of our concepts coincide with morphemes of a language, it follows that:

> (i) If you strip Mentalese of the word-type of which *T* is a token, the *only* other expressions which, in so doing, you'll be removing from Mentalese will be other morphemes,

and also that:

> (ii) Some of our concepts are such that one can have them without having others. This means that, if CTM is right, your concept of snow is *primitive*, the same being true of your concept of Smith, Jones, and Brown (assuming, for argument's sake, that the Mentalese expressions for these individuals are morphemic).

5.1 CTM's reliance on the erroneous assumption that there are minimal units of thought (continued)

But there are no primitive concepts. I couldn't possibly have a concept of snow, or of Barack Obama, or (to take a more tendentious example) of the number three without having a network of other concepts---extension, spatiotemporal location, surface-area, sentience, etc.

Concepts are necessarily contrastive; in order to have a concept of a thing, one has to be able to distinguish that thing from other things. A prerequisite for distinguishing a given rock from a given tree is my being aware of some property had by the rock that isn't had by the tree. And a prerequisite for my having such knowledge is my *already* having concepts.

5.1.1 How some contemporary philosophers would response to this argument

Many contemporary theorists---some of them CTM-advocates, some of them not—reject this line of thought, saying that, to have a "fix," on the rock (or the tree), I need only be on the receiving end of a causal series beginning with it. Their view is that concepts are mental photographs. The relationship borne by my concept of Socrates with respect to Socrates himself is identical with the relationship borne by a photograph of Barack Obama with respect to the events (the impact of certain light-rays with Barack Obama's body) that, in due course, led to that photograph.

The position just described involves a failure to distinguish between:

(i) awareness of some entity,

and:

(ii) information that, *if known*, would *lead* to such awareness.

I see discolorations of a certain kind on a photographic plate. Given that information, along with other pieces of background information that I have, I can make various inferences. I can infer that, at some point, light bounced off of a person who was bald and was wearing a toga (and so on). So the photographic plate is information in the sense that, to a creature with the requisite background and the requisite intelligence, its existence warrants the making of an inference that will eventuate in knowledge of some entity (e.g., a person) or state of affairs (e.g., one consisting of a person wearing a toga) that might otherwise have remained unknown. (Why "*might* otherwise have been unknown," instead of "*would* otherwise have been unknown"? Because I might have been the one who took the picture and might therefore have already known of the person and the corresponding state of affairs.) The discolorations on the photographic plate are "information" in the sense that, to a suitably endowed creature, they *warrant* the making of inferences that, in due course, lead to concepts.

In any case, it is massively to CTM's discredit that it presupposes that there are representational atoms.

5.2 A similar argument

Sentences are digital structures. Setting aside cases of ambiguity, any given sentence—any given complex expression, in fact---has a unique

decomposition into minimal units of significance. This means, as previously pointed out, that our thoughts decompose into minimal representational units. Which is radically implausible.

In any case, sense-perceptions don't have a digital structure, and neither, it would seem to follow, does the information that sense-perceptions encode. Given which, it would be very artificial to maintain that all post-perceptual cognition was digital or, therefore, sentential in nature.

But, for argument's sake, let's suppose that the contents of our sense-perceptions can, as some advocates of CTM maintain, be "digitized." Supposing this hypothesis correct, all the problems previously posed for CTM by perceptual content are now posed by the post-perceptual thoughts involved in this (alleged) digitization-process. If I'm translating from Albanian to English, I must speak both Albanian and English. Moreover, supposing that I'm translating competently, I have some way of evaluating the degree to which my English outputs match my Albanian inputs. If my thinking is entirely mediated by English sentences, I won't be able to do either of these things: I won't be able to translate from Albanian into English; and, if *per impossibile* I were able to do so, I couldn't verify the accuracy of my translations. By analogous reasoning, if the aforementioned post-perceptual thoughts were themselves entirely digital, they would *ipso facto* be blind to the benchmarks relative to which their own translations were to be evaluated, supposing for argument's sake that they could produce such translations to begin with.

Of course, one could say that those processes are purely causal; i.e., that they don't embody any awareness of what, given the relevant perceptual inputs, the relevant digital outputs *ought* to be. But in that case, our post-perceptual thoughts are *ex hypothesi* merely the *results of*, and are in no way *justified by*, the contents of our perceptions. But that is surely too absurd a

position to take seriously.

6.0 The regress-problem

If CTM is right, then:

(ARS)[339] anything that represents anything is a symbol.

If, as CTM says, our thoughts are sentence-tokens, then surely we understand those tokens. But understanding a sentence involves being able to pair it off with its meaning. This means that one must be able to compare that sentence with something that isn't a sentence. This wouldn't be possible if all of our thoughts were sentences (or sentence-tokens).

"That doesn't follow," it might be said. "Suppose that you think that snow is white, and that your thinking this consists in there being an occurrence of some Mentalese sentence. Your understanding that sentence could consist in there occurring, in your mind-brain, various sentences of some language—call it Mentalese_2---such that, given those sentence-occurrences, you can duly interpret the original (Mentalese) sentence."

Yes, but by that logic, there must be tokens of Mentalese_3 sentences before one can understand those Mentalese_2 sentence-occurrences; and there must be tokens of Mentalese_4 sentences before one can understand those Mentalese_3 sentence-occurrences; and so on *ad infinitum.*

So it's viciously regressive to hold, as CTM does, that symbol-tokens mediate our thoughts. More precisely, it's viciously regressive *given* the reasonable assumption that we *understand* those symbols.

6.1 How advocates of CTM deal with the regress-problem

Aware of this line of thought, advocates of CTM say that we *don't* understand the symbol-tokens that mediate our own thoughts. As William Lycan put it, we are simply "built to use those symbols" in the "right" ways, much as computers are "built" to use the symbols (or, rather, the things which are tendentiously and, we have seen, wrongly *called* "symbols") on which it operates.

6.2 The problem with the way advocates of CTM deal with the regress-problem

But, so far as a given thing isn't understood, it isn't functioning *as* a symbol. A story, similar to one told earlier, will make this clear. On my desk, I have a wooden sculpture, which we'll call "WS," that tokens the sentence "haste makes waste." Even though WS tokens three words, all three word-tokens are carved from the same block of wood. So WS consists of one, solid object, not of a lot of letter- or word-blocks. WS is heavy; it's made of solid oak, not pine or plywood or any other cheap sort of wood. It would thus serve as an excellent cudgel. A student of mine comes into my office. He reads it, understands it, and accordingly changes his, hitherto impetuous ways. In this context, WS is functioning as a symbol. In other words, if $E_1 \ldots E_n$ are the just-mentioned effects on the student, it is WS's being a symbol (of quite a specific kind, in fact) that leads to $E_1 \ldots E_n$. Later that afternoon, a colleague, whom I detest, comes into my office. He doesn't read WS. Because of his myopia, he sees a brown blur in the place that WS occupies. Gripped by a non-tenure-conducive rage, I strike my colleague with WS. If $E_1{*} \ldots E_m{*}$ are the subsequent bruises and fractures suffered by my colleague, it is WS's having a certain mass, moving with a certain velocity, etc. that brings about

E_1*….E_m*. WS's *being a symbol* is utterly irrelevant. Why? Because WS's being understood is essential to its functioning as a symbol.

Let us sum up. Languages *are* sets of semantic rules. Where there are no such rules, there are no languages; and where there is no knowledge of such rules, there is no linguistic activity. The idea that there is a "language of thought" is incoherent, given that it is inconsistent with these basic truths, and so is CTM, given that it assumes that there is a language of thought.

7.0 CTM self-refuting.

CTM says that we think in sentence-tokens. CTM also says that, *qua* proposition-bearing entities, those sentence-tokens *do nothing*. For reasons that we'll now see, it follows that, if CTM is correct, it cannot be *known* to be correct or even rationally accepted.

We must start by remembering that no *thing* is a cause. The rock isn't what breaks the window. What does so is the rock's striking the window with a certain amount of force. Bearing this in mind, let *B* be the brain-structure that encodes my belief that 1 + 1 = 2. Nobody denies that *B* has causal properties. But if CTM is right, *B*'s being (or constituting) a thought to the effect that 1 + 1 = 2 does nothing. The same thing *mutatis mutandis* holds of any content-bearing brain-state. If CTM is right, even if a given brain-state bears content *C*, its bearing that content is causally inert. This entails that, if a given person accepts CTM, it isn't because of the contents of any of his thoughts that he does so; and, consequently, it isn't because he knows of any reason to accept CTM. So, if CTM is correct, there is *ipso facto* no way that a person could be justified in accepting CTM (or, by obvious extensions of this reasoning, anything else). Which, while not showing that it's false, does show that one cannot coherently advocate it.

Hempel's paradox solved

Abstract: On at least one natural reading of it, (i) "if it's metal and it's being heated, it expands" is equivalent with(ii) "if it doesn't expand when heated, it isn't metal." Equivalence presumably entails confirmational equivalence: two statements are equivalent iff they're true in the same possible worlds. And given two such statements, a given body of data couldn't provide more support the one than the other without there *ipso facto* being worlds in which the one was true but the other was not. And yet (i) and (ii) seem not to be equivalent confirmationally, since, unless our intuitions are off, the existence of a non-expanding (non-metal) carpet fibers supports (ii) more than it supports (i). This paper argues that this problem can be solved without doing anything untoward---without contending *ad hoc* that confirmational and logical equivalence can pull apart, without claiming that our intuitions about confirmation are prejudiced (because, unbeknownst to us, hewed to illicit presuppositions), without engaging in the intricacies of formal probability theory, and without otherwise doing anything other than registering some pedestrian facts about the way in which understood meaning (pragmatics) diverges from literal meaning (semantics).

1.0 An illustration of the brokenness of Hume-based conceptions of causation and explanation: The raven paradox

It seems incontrovertible that:

(RP) Logically equivalent sentences are confirmationally equivalent.

This means that, if P and Q are logically equivalent, then not only does evidence E confirm P just in case E confirms Q, but E confirms P to a given degree just in case it confirms Q to that same degree.

For the record, the present author accepts RP. So did C.G. Hempel[340], who discovered that RP, if true, has, or seems to have, a paradoxical consequence.[341] The sentence:

(1) “All ravens are black”

is logically equivalent with:

(2) “All non-black things are non-ravens.”

The existence of non-ravens (e.g., pianos, pants, carpet fibers) that are pink or green confirms (2), but does *not*, unless our intuitions are very wrong, confirm (1). In any case, discovery of a black raven does more in the way of confirming (1) than does discovery of a pink piano, and discovery of a pink piano does more in the way of confirming (1) than does discovery of a black raven. So, impossible though it must be, (1) and (2) are logically, but not confirmationally, equivalent.

An analogous paradox is associated with each one of infinitely many other pairs of logically equivalent sentences, many of them of scientifically important. For example:

(i) “if it’s made of metal and is being heated, then it’s expanding.”

is logically equivalent with:

(ii) “if it isn’t expanding, then it isn’t both made of metal and being heated.”

(ii) is confirmed by the fact that the non-expanding Tupperware in my freezer is not made of metal (or *a fortiori* of metal that is being heated). But, unless our intuitions are very wrong, that fact provides no support for (i), or at least provides less support for it than it does for (ii).

Some, including Hempel himself, have argued that, our intuitions notwithstanding, the existence of pink pianos *does* confirm (1) to degree n exactly if it confirms (2) to that same degree and that the existence of non-expanding Tupperware in freezers does confirm (i) exactly as much as it confirms (ii). Others (e.g. Quine[342]) have argued that the paradox should be solved by advocating non-standard logics, relative to which logical equivalence doesn’t guarantee confirmational equivalence.[343]

Whatever merits these solutions have, it’s worth trying to solve Hempel’s paradox in a manner congenial to commonsense-friendly and also with classical (meta)logical principles. And we’ll now see that such a solution is available.[344]

The trick is to look at what (i) and (ii) *convey*, and not just at what they can (reasonably) be taken to have for their literal meanings. The trick, in other words, is too look, not just at semantics, but also at pragmatics. Bearing this in mind, let us now (in Section 1.1) discuss an assumption we’ll make concerning the semantics of categoricals.[345]

2.0 Inductions by enumeration identical with crypto-inferences to the best explanation

In this context, (i) may be treated as saying the same thing as:

> (MH) For any x, if x is made of metal and x is being heated, then x is expanding.

And (ii) may be treated as saying the same thing as:

> (HM) For any x, if x isn't expanding, then x isn't made of metal that is being heated.

MH can be taken *either* to express an "accidental generalization" *or* to express a principled relationship. Taken in the first way, it does nothing more than affirm a purely *de facto* concomitance of two properties (viz. the property of being heated metal and the property of being something that expands). Taken in the second way, MH seems to say that all heated metal expands, *the reason being* that a thing's being heated metal provides some kind of nomic basis for its expanding. In this section, we will focus on the first, strictly extensional, reading. In the next, we will focus on the second, nomically committed reading.

Let us paraphrase MH, when it is taken as a purely *de facto* generalization, as follows:

> (MHA)[346] For any x, if x is made of metal and is being heated, then, by happenstance, x is expanding.

So, in this context, take the operator “by happenstance” to be short for “not because x’s having the first property provides a causal or nomic basis for its having the second.” (By “the first property,” I mean the property mentioned before the “then,” viz. the property of being made of metal that is being heated; and by “the second property,” I mean the property referred to after the “then,” viz. the property of expanding.) We don’t need to worry about whether this corresponds to the usual meaning of “as a matter of happenstance,” since we are using this expression in a stipulative manner.

If we take MH in this way—if we take it to mean the same thing as MHA—does Hempel’s paradox arise? No. Let phi and psi be any two properties such that *x has psi* isn’t an analytic consequence of *x has phi*, or *vice versa*. If I know that there is no causal or, more generally, nomic connection between x’s being a phi, on the one hand, and x’s being a psi, on the other, then I have no reason to infer, from x’s being a phi, that x is also a psi. Even if, thus far, every ball withdrawn from the urn has been black, unless I have reason to believe that this concomitance has some basis in natural law or in some causal mechanism (e.g., the presence of a force field surrounding the urn that allows entry only to black balls, or some person’s resolute intention to allow no non-black balls into the urn), I must *ipso facto* regard it as nothing but coincidence that, thus far, every ball withdrawn has been black; and I must therefore *not* think it more likely than not that the next ball withdrawn will be black—for to think otherwise would be to commit the Monte Carlo fallacy. [347] Thus, supposing that, in fact, this concomitance is mere coincidence, the generalization *all the balls in the urn are black* fails to receive any positive confirmation from its instances. [348] For the same reason *mutatis mutandis*, the statement *all non-black things are not balls in the urn*, unless read as having nomic force, fails to receive support from *its* instances.

And, of course, it also follows that HM, if read as doing nothing more than

affirming a *de facto* concomitance, doesn't receive support from its instances. Thus

> (HMA) For any x, if x isn't expanding, then, by happenstance, x isn't made of metal that is being heated,

doesn't receive support from the fact that the non-expanding Tupperware in my freezer isn't made of metal that is being heated.

So, if MH and HM are read as having no nomic content—if they are read as doing nothing but expressing *de facto* concomitances—then, once it is seen that *de facto* generalizations don't receive support from their instances, HM doesn't even *appear* to receive support from the fact the non-expanding Tupperware in my freezer isn't made of heated metal, and MH doesn't even *appear* to receive support from the fact that the metal pan I put in the oven is expanding. A fortiori the former fact doesn't even *appear* to confirm HM more than it does MH, and the latter fact doesn't even appear to confirm MH more than it does HM.

And we therefore don't get the kind of (apparent) confirmational non-equivalence that gives rise to Hempel's paradox. For that paradox depends precisely on its being the case that, despite the (alleged) fact that HM and MH are logically equivalent, they are not, or at least appear not to be, confirmationally equivalent. And they appear to be confirmationally non-equivalent only in so far as (e.g.) the existence of a heated, expanding metal objects confirms MH but not HM, and only in so far as (e.g.) the existence of non-expanding, non-heated, non-metal objects confirms HM but not MH. But once it is made clear that *de facto* generalizations don't receive support from their instances, it follows that, if MH and HM are read as free of nomic content, nothing even appears to give more support to the one than to the

other. And the paradox fails to arise.

2.1 Inductions by enumeration identical with crypto-inferences to the best explanation (continued)

So, if Hempel's paradox is to arise, at least one of MH and HM must be taken as doing more than affirming a *de facto* concomitance. And it seems that, thus taken, MH is to the effect that all heated metal expands *for the reason that* a thing's being metal that is being heated provides some kind of nomic or causal basis for its expanding. Let us paraphrase MH, when it is taken in this way, as follows:

> (MHN)[349] For any x, if x is made of metal and is being heated, then, as a matter of principle, x is expanding.

So, in this context, take the operator "as a matter of principle" to be short for: "because x's having the first property provides a causal or nomic basis for its having the second." Echoing what we said earlier, it is irrelevant this may not correspond very well to the usual meaning of "as a matter of principle," since we are here using this expression in a stipulative manner.

At this point, we have a choice. We can take HM as having no nomic content—as taking it to have the same import HMA. Or we can take it to have nomic content. Let's consider both options, starting with the first.

Given that MH is taken to do more than affirm a *de facto* concomitance, if MH is taken to affirm nothing but a *de facto* concomitance, the two will, quite trivially, not be equivalent. For obviously MHN is not equivalent with HMA. The first is committed to its being the case that *inter alia* a thing's being made of heated metal provides some kind of nomic basis for—or, in

any case, is nomically or causally connected in some way with—that thing's expanding. The second isn't committed to there being any kind of nomic, or otherwise not purely extensional, connection between the property of not expanding and the property of not being made of heated metal. MHN is true only if certain nomic connections hold; HMA can be true in the absence of those connections. So there is no equivalence, or even the appearance of equivalence for that matter.

We've just seen that, if we take MH to have nomic force (which we must—see above—if Hempel's paradox is to arise), Hempel's paradox doesn't arise if we *don't* take HM to have nomic content. In other words, if we take HM to do nothing more than affirm a *de facto* concomitance, then it won't, as we've just seen, even appear to be equivalent with MHN, and nothing even resembling Hempel's paradox arises. So if that paradox is to arise, then we must take HM to have nomic content.

In light of this, consider the sentence:

> (HMN) For any x, if x isn't expanding, then, as a matter of principle, x isn't made of metal that is being heated,

HMN says that a thing's not being an expander provides some kind of (partial) nomic basis for its not being heated metal. Of course, the relationship that HMN bears to HM is analogous to the relationship that MHN bears to MH. But what is most important here is that HMN would be the right way—or, in any case, the (or, at least, *a*) natural way—to take HM *if* the latter were taken to do more than affirm a *de facto* concomitance between two properties.

To clarify this last point, suppose that you are speaking to a friend who is about to drink a large quantity of arsenic but who doesn't realize that

consuming arsenic in large quantities leads to death. You tell him, warningly, "people who consume large quantities of arsenic die soon thereafter." Here you are obviously not affirming a purely extensional connection between, on the one hand, the property of being a person who drinks large quantities of arsenic at a given time and, on the other hand, the property of being something that dies soon thereafter. What you are saying is that a person x's drinking large quantities of arsenic at a given time *leads to,* or in any case provides some kind of nomic foundation for, x's dying soon thereafter, *that* being why there is a concomitance between the property of being a person who drinks arsenic, on the one hand, and the property of being dead soon thereafter, on the other. That this is the right interpretation is made clear by the fact that you would be *failing* to warn your friend not to drink the arsenic if you had instead said: "all people who have drunk arsenic died soon thereafter. But that had nothing to do with their drinking arsenic. You see, it turns out that the only people who've ever drunk arsenic belong to religious sect X; moreover, all of those people were about to be executed. One of the precepts of X is that, if you are about to be executed, you must drink arsenic in order to get into heaven. And that's why, thus far, each person who's consumed arsenic died soon thereafter.

Just as the most natural way to take "people who drink arsenic die"/"heated metal expands" to do more than affirm a *de facto* concomitance is to take it to mean that a person's drinking arsenic/being heated metal provides some kind of causal or nomic basis for that person's dying/expanding, so the most natural way to take HM to have nomic force is to take it to mean that a thing's being a non-expander provides some kind of causal or nomic basis for its not being heated metal. And that is precisely the reading of HM that is embodied in HMN.

HMN is clearly *not* equivalent with MHN. Suppose, if only for argument's

sake, that a thing's being made of heated metal *causes* it to expand. So suppose that:

> (MHC) "as a matter of natural law, x's being heated metal *causes* x to expand."

That would certainly be sufficient for the truth of MHN. In light of this, consider the statement that:

> (HMC) "as a matter of natural law, x's not expanding causes x to not be metal that is being heated."

Suppose it is a law that *x's being a human who drinks arsenic at t causes x to be dead at some later time t**. This very reasonable and (give or take a few nuances) correct statement obviously doesn't entail, and therefore isn't equivalent with, the false and incoherent statement that *x's not being dead t** *causes x to be a human who doesn't drink arsenic at earlier time t.*

For similar reasons, MHC is not equivalent with HMC. So, since MHC and HMC are sufficient for the truth of, respectively, MHN and HMN, it follows that MHN and HMN are not equivalent: they don't hold in the same models, the same possible worlds. In fact, whereas MHN, or something similar to it, *does* hold in our world, the same doesn't go for HMN or anything at all like it.

2.2 Inductions by enumeration identical with crypto-inferences to the best explanation (continued)

We've seen, first, that MH must be read as having some kind of nomic content if Hempel's paradox is to arise (i.e., if there is to even be the appearance that MH is logically, but not confirmationally, equivalent with HM). And we've also asserted that, if MH read as having nomic force, its meaning is given by MHN, and we made the same assumption *mutatis mutandis* about HM.

But, it must be admitted, we have provided no argumentative basis for this assertion, our grounds for it being purely intuitive. Which brings us to our next point, namely, that there is another possible nomic reading of MH:

> (MHW[350]) "thanks to the existence of some kind of nomic (or causal) connection between the two properties about to be mentioned [viz. the property of being heated metal, on the one hand, and the property of expanding, on the other], for all x, if x is heated metal, then x is expanding."

How is MHN different from MHW? MHN says pretty much the same thing as:

> (*) for all x, if x is heated metal, then x is expanding, and this is because x's being heated metal provides a nomic basis for x's expanding.

By contrast, MHW says that:

> (^) for all x, if x is heated metal, then x is expanding, and this is because *either* x's being heated metal provides a nomic basis for x's expanding *or* x's expanding provides a nomic basis for

x’s being heated metal.

To be sure, the expression “provides a nomic basis for” is an obscure one, which I’ve done little or nothing to clarify; and it might therefore seem that there is no difference between (*) and (^). But, obscure though that expression may be, (*) and (^) are demonstrably different, and non-equivalent, statements.

It’s true (or so let us suppose, for argument’s sake) that all people who consume large quantities of arsenic die soon thereafter. And this is not an accident: human biology being what it is, the chemical properties of arsenic being what they are, there’s a nomic connection between consuming arsenic and dying. Vague as the expression “provides a nomic basis for” is, would it be correct to say that being a dead human at 3:10 P.M. “provides a nomic basis for” being a person who drinks arsenic at 3:00 P.M.? More generally, supposing that x’s having psi at t* is an *effect* of x’s having phi at earlier time t, does it make sense to say that x’s having psi at t* provides a nomic basis for x’s having phi at t? No. My being obese provides a nomic basis for my having a heart-attack. But surely my having a heart attack doesn’t provide a nomic basis for my being obese. And that’s
all we need to know in this context about what it means to say that one thing provides a “nomic basis” for another. So let us proceed.

Suppose that, in possible world W, all things made of heated metal expand. Further suppose that, in W, x’s being something that is expanding causes x to turn into something that is metal whose temperature is continuously rising and whose volume is constantly increasing. And, to make things simple, suppose that, in W, all things made of heated metal got that way by, at some point in their histories, being things that were expanding. In that case, MHW would be correct in W: for there would, in that world, be a

nomic connection between a thing's being heated metal, on the one hand, and its being something that expands, on the other; moreover, the fact that, in W, all heated metal expands would have a basis in this nomic connection. But it would not, at least not on that account, be correct to say that, in W, a thing's being made of heated metal provides a nomic basis for its expanding. In W, things expand *before* they become expanding heated metal. So, in W, a thing's being heated metal obviously cannot *cause* it to expand.

But there is a further point: in W, a thing's expanding is not to be *understood in terms of its being heated metal.* It is the other way around. In W, a thing's expanding is causally *and explanatorily* more basic than its being heated metal. In any case, just as (see above) it would be wrong, or at least very odd, to say that being dead at t* provided a *nomic basis* for drinking arsenic at earlier time t, it would, for parallel reasons, be wrong to say that, in W, being heated metal provided a nomic basis for being a thing that expands. And this entails that, however vague the expression "provides a nomic basis for" is, * and ^ are, very clearly, *not* equivalent.

3.0 This argument concluded

First of all, MHW obviously isn't logically equivalent with HMA: the former does, whereas the latter does not, allege the existence of certain nomic dependencies. More importantly, MHW isn't equivalent with:

> (HMW) "thanks to the existence of some kind of nomic (or causal) connection between the two properties about to be mentioned [viz. the property of not expanding, on the one hand, and the property of not being heated metal, on the other], for all x, if x is not expanding, then x is not heated metal."

Given only that there is a nomic connection between a thing's having phi and its having psi, it doesn't follow that there is a nomic connection between a thing's *not* having psi and its *not* having phi. There is a causal, and therefore a nomic, connection between Smith's falling off of tall building X at t, on the one hand, and his being dead at later time t*, on the other. Smith's falling off of X is an event. It is something that makes things happen (e.g., various bones of Smith's being shattered); and, what is probably a more general notion, it is something in terms of which things can be explained.

Compare Smith's situation with Brown's. Brown does *not* fall off X, or any other building, at t; and Brown is alive and well at t*. Obviously Brown's *not* falling off X at t doesn't *cause* Brown to be alive at t*. More generally, Brown's being alive at t* is not, except perhaps in some extremely derivative sense, to be explained in terms of Brown's not falling off of X. Whereas Smith's falling off of X is an event, Brown's *not* falling off of X is *not* an event, and it doesn't consist of events. A chair consists of events (various miniscule displacements of mass-energy), and so does a corporation (various actions on the part of sentient beings); and so, indeed, does any enduring entity—any structure, any "object," any non-instantaneous event. But Brown's *not* falling off X isn't identical with an event and isn't constituted by events.

Of course, since it's a truth, and therefore a "fact," that Brown did not fall off of X, we might say that the (negative) "fact" of his *not* falling off of X exists no less than the (positive) fact of Smith's falling off of X. But that doesn't undermine our point that Brown's *not* falling off of X isn't identical with, or constituted by, events.

While it is a matter of intense debate what nomic relations are, it is pretty clear they hold only among occupants of the space-time manifold. They

clearly hold among events (my pounding the table causes it to break). Maybe they hold among things other than events. Maybe, for example, they hold among conditions. And maybe they hold among objects. But while it is unclear whether *all* spatiotemporal entities stand in nomic relations, it *is* clear that nothing *non*-spatiotemporal stands in them. My pounding the table is nomically connected to the table's breaking. But my *not* destroying Jupiter isn't nomically related to anything.

My pounding the table is an event; it makes things happen; things can be explained in terms of it; it has a place in the tapestry that constitutes the spatiotemporal world. My *not* destroying Jupiter is nowhere in that tapestry, and therefore isn't fit to be nomically related to anything. There is no event of my not destroying Jupiter. There is thus no condition or object that is constituted to any degree by that non-event. Therefore, it isn't identical with, or constitutive of, anything that falls within the scope of natural law. It isn't, therefore, nomically connected to anything.

A thing's being composed of metal that is of a certain temperature is, in the relevant respects, like my pounding the table. Things are to be explained in terms of it. And, though not itself an "event," it is *constituted* by various interacting events, all of which occupy space-time. So the skillet's having a certain temperature, at a given time, is something that can, and indeed obviously must, stand in nomic relations. And the same is true of the skillet's expanding during a given interval.

But a thing's *not* expanding is, in the relevant respects, like my *not* destroying Jupiter, and so is a thing's *not* being heated metal. So, whereas a thing's being heated metal is, or at least could be, nomically related to its expanding, it doesn't follow, and it isn't the case, that a thing's *not* expanding could be nomically related to its *not* being heated metal.

These points make it clear why MHW isn't logically equivalent with

HMW. A thing's being heated metal *can* (and, in fact, must) be nomically related to other things, and the same is true of a thing's expanding. But a thing's *not* expanding, like my *not* destroying Jupiter, isn't nomically related to anything, and neither is a thing's *not* being heated metal. And that's why, even though, when read as merely affirming *de facto concomitances*, MH and HM are equivalent, they cannot always be intersubstituted *salva veritate*. In particular, they cannot be so intersubstituted in contexts governed by the operator that begins each of MHW and HMW (viz. "'thanks to the existence of some kind of nomic (or causal) connection between the two properties about to be mentioned"). There is nothing mysterious about this. Many operators don't permit intersubstitutions of logically equivalent expressions. Logically equivalent sentences can be truth-preservingly intersubstituted in contexts governed by (e.g.) "it is true that," but not, at least not categorically, in contexts governed by "Timmy believes that." The operator "there is a nomic connection between the facts expressed by the antecedent and consequent of the following statement" doesn't tolerate intersubstitutions of logically equivalent sentences. For, even if:

> (T_1) "there is a nomic connection between the facts expressed by the antecedent and consequent of the following statement: if x is heated metal, then x is expanding"

is true, it doesn't follow that

> (T_2) "there is a nomic connection between the facts expressed by the antecedent and consequent of the following statement: if x isn't expanding, then x isn't heated metal"

is also true. In fact, T_2 is, for the reasons just given, quite plainly false.[351]

Summary and Conclusion

If read as doing nothing more than affirming *de facto* concomitances:

(MH) if x is made of heated metal, x expands

and

(HM) if x isn't expanding, x isn't made of heated metal

don't receive support—and, to anyone who isn't guilty of the gambler's fallacy, don't even appear to receive support—from their instances and are therefore vacuously confirmationally identical. So, thus read, they are logically *and*, quite clearly, confirmationally identical; and Hempel's paradox doesn't arise. If they are read as doing *more* than affirming *de facto* concomitances, they must be read as alleging, or at least being committed to there being, some kind of nomic connection between the properties in question. But, no matter what nomic readings we give these sentences, they don't even appear logically equivalent. (This is, to put it very crudely, because logically equivalent sentences cannot, at least not in general, be intersubstituted *salva veritate* in contexts governed by nomic operators; e.g. "there is a nomic connection between the following statements.") Bottom line: once it is realized that accidental generalizations don't receive support from their instances, this being the prerogative of principled generalizations, it isn't possible for Hempel's paradox even to arise.

References:

Bach, Kent. 1984. Language and Thought. Oxford: Clarendon.
Bunch, Bryan. (1981). Mathematical Fallacies and Paradoxes. Mineoloa, New York: Dover.
Carnap, Rudolph. (1953). “Two Concepts of Probability.” In Feigle, H. and Broderick, M. eds. Readings in the Philosophy of Science. New York: Century-Crofts-Appleton pp. 438-455.
Davis, Wayne. 2003. Meaning, Expression, and Thought. Cambridge: Cambridge University Press.
Dupré, John (2000). “Paradoxes of Confirmation.” In Newton-Smith, W.H. (ed.). A Companion to the Philosophy of Science. New York: Blackwell, pp. 402-404.
Hempel, Carl. G. (1945) *Studies in Logic and Confirmation.* Mind 54, 1-26.
Hempel, Carl. G. (1945) *Studies in Logic and Confirmation. II.* Mind 54, 97-121.
Hempel, Carl G, *Studies in the Logic of Confirmation.* In Marguerite H. Foster and Michael .
Hempel, Carl. G. (1965). “Empiricist Criteria of Cognitive Significance.”
L. Martin eds. (1966) *Probability, Confirmation, and Simplicity*. New York: Odyssey Press, pp. 145-183
Falletta, Nicholas. (1983) *The Paradoxicon: a Collection of Contradictory Challenges, Problematical Puzzles, and Impossible Illustrations.*
Cohen, Jonathan (1989) *The Philosophy of Induction and Probability*, Clarendon Press - Oxford.
Gardner, Martin (1973) *Mathematical Puzzles and Diversions*, Penguin Books.

Leavitt, F. (1996). “Resolving Hempel’s Raven Paradox.” Philosophical Inquiry 18, nos 3/4: 116.
Lange, Marc. (2007) *The Philosophy of Science.* Oxford: Blackwell.
Maher, P. (1999). “Inductive Logic and the Raven Paradox”. *Philosophy of Science* 66: 50-70.
Quine, W.V.O. 1977. Ontological Relativity. New York: Columbia University Press.
Percival, Philip (2000). “Paradoxes of Confirmation.” In Newton-Smith, W.H. (ed.). A Companion to the Philosophy of Science. New York: Blackwell, pp. 358-372.
Ray, Christopher (2000). “Paradoxes of Confirmation.” In Newton-Smith, W.H. (ed.). A Companion to the Philosophy of Science. New York: Blackwell, pp. 243-251.
Sainsbury R. M. (1997) *Paradoxes*, (2nd Ed.) Cambridge University Press.
Scheffler, Israel (1981) *The Anatomy of Inquiry.* Indianapolis: Hackett.
Schlesinger, G (1974) *Confirmation and Confirmability*, Clarendon Press - Oxford.
Trout, J.D. (2000). “Paradoxes of Confirmation.” In Newton-Smith, W.H. (ed.). A Companion to the Philosophy of Science. New York: Blackwell, pp. 53-55.
Trusted, Jennifer (1979) *The Logic of Scientific Inference*, The MacMillan Press Ltd.

What is Formal Truth?

Abstract: A formal truth is one that is an instance of a sentence-schema all of whose instances are true. Other analyses of formal truth are shown to be viciously circular, extensionally false, or both. Among these alternative analyses are the contentions that that formal truth is identical with syntactic truth; with truth under all reinterpretations of non-logical terms (Quine (1970)); with maximally generic truth (Russell (1921); and with topic neutral truth (Frege (1884)). It is shown that not all logical truths are formal truths. This last fact has been known for a while; for example, Gödel (1953) showed that followed from his own work on incompleteness, and, in a grossly undervalued work, Arthur Pap (1958) established much the same position, or came close to doing so, as did Quine (1936) in *Truth by Convention*.[352] But our particular way of establishing it will help bring to light its long-overlooked bearing on the nature of formal truth. Also, in (re)establishing this result, we'll establish one that, in our own time, is widely seen as false, viz. that on any coherent delineation of the term "logic," logic cannot be identified with the study of formal truth, a consequence being that the most fundamental truths of logic---be it the "logic" of Alonzo Church or that of F.H. Bradely---must consist of non-formal, analytic truths.

1.0 Formal truth as syntactic truth Here are some sentences[353] that would generally be described as "formally true" (in this paper, quotation marks are, whenever the context requires it, to be treated as quasi-quotation marks):

(1) “If Bill is tall, then it is not the case that Bill is not tall.”
(2) “If everything has a given property, then any given thing has that property.”
(3) “If all F’s are G’s, and no G’s are H’s, then no F’s are H’s.”
(4) “Either Helsinki is in Finland or Helsinki is not in Finland.” [354]

What is “formal truth”? What does it mean to describe (1), for example, as “formally true”? Here’s the canonical answer:

(FT) Formal truth is *syntactic* truth. A sentence is “syntactically true” if, given only its syntactic structure, it cannot fail to be true. Thus, a sentence is formally true if no sentence having the same syntactic structure is false.

FT cannot be accepted. A linguist would say that (1) “grass is green or grass is not green” has the same syntax as (2) “grass is green and grass is not green.” As linguists use the word “syntax,” two expressions have the same syntax if the way in which the one is built out of its parts is identical with the way in which the other is built out of *its* parts. For example,

(#) “Bill hates Ted”

and

(##) “Laura loves Sally,”

though having very different meanings, have the same *syntax*, the reason being that they're structurally identical. The one is built out of its simple constituents in the very way that the other is built out of *its* constituents. [355]

The relevant difference is *sub*structural. It lies in the fact that the *meaning* of "and" is different from that of "or." So the difference between (1) and (2) is like the difference between "John is tall" and "Mary is short": it lies entirely in the fact that the primitive, semantically unstructured constituents of the one have different meanings from their counterparts in the other. Thus, relative to what *linguists* mean by "syntax" (and "syntactic," etc.), it's a non-starter to say that "formal truth" is syntactic truth. Logicians and philosophers obviously have in mind some very different understanding of the word "syntax" when they identify "formal" truth with "syntactic" truth. They *don't* have in mind sentential (or phrasal) structure. What *do* they have in mind? What do philosophers mean when they say that a formally true sentence is one that holds entirely in virtue of its "syntax"? What do they mean when they describe some statement as a "formal truth"?

2.0 Quine on formal truth: logical/formal truth as truth under all reinterpretations of non-logical expressions W.V.O. Quine (1970) gave a characteristically intelligent and straightforward answer to this important question: a sentence S is formally true if, given any expression in S that isn't a *logical constant*, it isn't possible to produce a false sentence by replacing that expression with another expression of the appropriate grammatical

category.[356] A “logical constant” is an expression on the following list: “if . . . then . . . ,” “either . . . or . . . ,” “not,” “for all,” “for some,” “=”—and any obvious derivatives of any of these expressions (e.g., “is identical with,” “given any,” “it is not the case that,” etc.)[357]

Consider the sentence “grass is green or grass is not green.” Replacing “grass” with “fire” or “Bob” results in a truth, as does replacing “green” with “cool” or “voluminous.” (It’s assumed that the replacements are uniformly made: if a single occurrence of “grass” is replaced with “fire,” so is every other such occurrence.) Given any expression in (1) *other* than the logical constants in it (viz., “either . . . or . . .” and “not”), replacing that expression with another belonging to the same grammatical category results in truth.

There are two problems with Quine’s analysis. First, it doesn’t say what is so special about those items; it doesn’t say what these so-called “logical constants” have in common with one another that other expressions *not* on the list *don’t* have in common with them. And until that information is provided, there’s nothing to prevent one from extending that list in arbitrary ways. Unless that list was generated by some rule, anything can be added to it without diminishing its integrity, since it has no integrity to begin with. But if it was generated by some rule, then Quine’s answer fails to identify it and thus fails to say what formal truth is.

An analogy may help. When asked to define the expression “number,” I simply *list* the various different kinds of numbers (“real,” “whole,” “rational,” “imaginary,” etc.). Have I answered the question? No. Even though I’ve listed the various different kinds of numbers, I haven’t said what numbers are, and I thus haven’t provided any *reason* why fractions are on that list but birds are not. Accordingly, my answer doesn’t give anyone a reason not to extend that list as he or she pleases—not to decide that birds

and rocks are numbers. Of course, nobody *would* extend the list in such an absurd way. But that's only because our intuitions strongly suggest that there is *some* characteristic had in common by reals, rationals, etc., that birds and rocks don't have.

Second, there are truths not covered by Quine's analysis that don't appear to differ in any relevant respect from those that are covered by it. Many modal statements are in this category—for example, "if it's necessarily the case that Smith is a mammal if Smith is a dog, then if it's necessarily the case that Smith is a mammal, then it's necessarily the case that Smith is a dog." There doesn't seem to be any relevant difference between *that* statement and one that conforms to Quine's analysis—for example, "if Smith is a mammal if Smith is a dog, and Smith is an animal if Smith is mammal, then Smith is an animal if Smith is a dog." So there doesn't seem to be any good reason why "necessarily," unlike "if . . . then . . . ," doesn't make it onto Quine's list.

Of course, one could deal with this by adding "necessarily" to that list. But then we'd have to ask: "what makes *that* extension of the list legitimate?" –bringing us back to the criticism made a moment ago.

3.0 Formal truth as topic-neutral truth Some (e.g. Frege (1884), Russell (1920)) have argued that formal truth is maximally generic, or "topic-neutral," truth. The idea is that "$x = x$" is logically true because it's completely general: it doesn't belong to any one discipline and is presupposed by all of them.[358]

But this doesn't work. There are maximally generic truths that aren't formal—for example, "nothing can wholly pre-exist itself," "nothing spatiotemporal can lack causal properties," "nothing non-spatiotemporal can have causal properties," "inductive inference isn't an appropriate method of reasoning for non-empirical disciplines." Given any one of these sentences,

it's easy to show that there are false sentences having the same syntactic form. For example, the first of those sentences has the same form as "nothing spatiotemporal can have causal properties," which is obviously false.

One could argue that these truths *aren't* sufficiently generic. But then it would be very unclear what exactly "sufficiently generic" meant in this context, and the suspicion would arise that, so far as it had any coherent meaning, "sufficiently generic" meant "not a counterexample to this particular theory of formal truth."

Also, whether a sentence expresses a generic truth has nothing to do with its form. The truth expressed by "if everything has a given property, then any given thing has that property" is surely a maximally generic one. But that truth could obviously be expressed by a primitive, syntactically unstructured symbol. In fact, there is *no* form that sentences must have if they are to express that truth or any other. So, if "formal truth" is taken to be a characteristic of *expressions* of some kind, then it's simply false to identify "formal truth" with "maximally generic" truth. And if "formal truth" isn't meant in this way, then it's simply irrelevant to our current inquiry.

4.0 Formal truth as analytic truth The two main contentions of Wittgenstein's *Tractatus Logico-philosophicus* (TLP) are (i) that all non-empirical truths are tautologous (i.e., that they are definitional truths; e.g., "a fortnight is a period of two weeks") and (ii) that all non-tautologous truths are observation-reports.

Given the points just made, we are in a position to put forth a reason to reject (i) and therewith to reject a third contention of the TLP, viz., that all analytic truth is formal truth and, therefore, that all entailment is formal entailment.

The open sentence "If *P* and *Q* and *R*, then *R*" is true for all values of its

variables. This means that:

> (A) For any propositions *P*, *Q*, and *R*, ‹if *P* and *Q* and *R*, then *R*›[359] is true.

But, (A) is not itself a formal truth. (A) is true, of course. But it is not in virtue of its being an instance of a universally true open-sentence that it's true. For it isn't an instance of such a sentence. Any open-sentence of which (A) is an instance is one of which:

> (B) For no propositions *P*, *Q*, and *R*, ‹if *P* and *Q* and *R*, then *R*› is true,

which is obviously false (and thus not true, formally or otherwise). Since (B) is an instance of any open-sentence of which (A) is an instance, the former is not formally true.

It is in virtue not only of its form, but also of what is meant by "for any propositions *P*, *Q*, and *R*" that (A) is true. (A) and (B) have the same form. What distinguishes them is that, in the place where (B) contains an occurrence of one expression ("for *no* propositions *P*, *Q*, and *R*"), (A) contains an expression with a different meaning ("for *any* propositions *P*, *Q*, and *R*").

So the fact that A is analytic, whereas B is not, cannot possibly be chalked up to the fact that one is formally true, whereas the other is not. Thus, *not all analytic truth is formal truth.*

In the TLP, Wittgenstein[360] says that all analytic truth *is* formal truth. Wittgenstein is wrong, for the reasons just given. Also, even if there were

some formally true sentence *S* such that *S* was equivalent with (A), the statement (in this context you make take "statement" to mean "sentence," "sentence-token," or "proposition"):

(C) *S* is equivalent with A.

would not *itself* be a formal truth. To be a formal truth is to be a consequence of an informally true universal generalization. The sentence:

(%) if grass is green and snow is white and roses are red, then roses are red

is a formal truth because it is a consequence of (A). Thus, Wittgenstein's thesis that all analytic truth is formal truth is not only wrong but incoherent.

Here is another illustration of that fact. The sentence:

(JTB) any case of knowledge is a case of justified true belief,

is an analytic, not an empirical, truth. But it isn't a formal truth, since it has the same form as:

(NTB) any case of knowledge is a case of unjustified false belief,

which is false.

Wittgenstein grants that there *seem* to be analytically true sentences that aren't formally true. But he denies that there really are such sentences. His position is that, although it *seems* not to be formally true, the sentence:

(JTB) any case of knowledge is a case of justified true belief,

is in fact formally true, the reason being that it is an abbreviation of:

(JTBF) any case of non-accidentally true, justified belief is a case of justified true belief,

or, at any rate, of some other formally true sentence.

Contrary to what Wittgenstein alleges, JTB is not JTBF in disguise. They express different propositions. JTB is informative. JTBF is not. One can accept the proposition meant by the one without accepting the proposition meant by the other. JTB and JTBF express *equivalent* propositions (assuming, as we are, that JTBF is correct). But if JTB is to be a mere abbreviation of JTBF, both must express the very same proposition. They don't; so the view that analytic truth is the same thing as formal truth (henceforth "AFT") is false.

There is another problem with AFT. Any given informal analytic truth can be turned into a formal truth by replacing expressions with analytic definitions of themselves—by replacing "circle" with "closed planar figure of uniform curvature," etc. But such formalization would *presuppose* the correctness of all the analyses involved, and none of those analyses would be expressed by formal truths. JTBF is indeed formal. But it doesn't express an analysis. What *does* express an analysis (or so we will suppose for argument's sake) is:

(JTBA) any case of knowledge is a case of non-accidentally true, justified belief.

But JTBA isn't *formally* true, the reason being there are formally identical false statements (for example, "any case of knowledge is a case of accidentally true, unjustified belief").

4.1 Formal truth as analytic truth (continued) First a terminological point: The sentence "condition *C* is fulfilled" is, as we'll be using it, synonymous with "property *P* is instantiated."

A conceptual analysis is what is expressed by a non-empirical, non-trivial sentence of the form "C_1 is fulfilled if and only if C_2," where C_1 and C_2 are conditions. Thus:

> (CK) *x* is a circle if and only if *x* is a closed planar figure of uniform curvature,

is a conceptual analysis, and so is:

> (CK*) *x* is a closed planar figure of uniform curvature iff , for some plane *P* and some point D on *P*, *x* is the area bounded by every point on *P* that is a fixed distance from *D*.

Neither CK nor CK* is a formal truth.

No conceptual analysis is formally true. Conceptual analyses are expressed by analytic statements, and analytic statements either: (i) coincide with informal, analytic universal generalizations or (ii) are instances of such generalizations. If (i), they're obviously not formally true. But any statement that is an instance of some informal, analytic universal generalization is trivial, and is therefore not analytic. For example, % is a consequence of (A),

and % is clearly trivial.

4.1.1 A similar argument The following sentence is formally true:

> (i) “snow is white only if snow is white or grass is green”

But the following is *not* formally true:

> (ii) “it’s formally true that ‘snow is white only if snow is white or grass is green’ is formally true.”

It’s an empirical truth, not a truth of logic, that (i) is even an expression, let alone a true, or *a fortiori* formally, true one. To be sure, it is *not* an empirical truth that, the semantic and syntactic rules of English being what they are, (iii) is true. In other words,

> (iii) Assuming that $S_1 \ldots S_n$ are the semantic/syntactic rules of English, it (logically) follows that is that (i) is formally true.

But (iii) is not itself a formal truth. But there’s no denying that it’s a truth of logic. But for (iii), or some other truth of its ilk (viz. some non-empirical, informal truth), there’s no way to establish that (i) is formally true. In general, informal truths are needed to establish whether a given sentence is formally true or not. It’s only on the basis of informal truths that it can be determined whether a given truth is formal or not. In other words, *if* formal truth is logical truth, then it’s only on the basis of non-logical truths that it can be determined whether a given statement is formally true or not---a statement’s

being logically true would be a non-logical consequence of some non-logical truth. But in that case, the term “logic” wouldn’t refer to the body of truths with respect to which the legitimacy of an inference or argument is to be evaluated; it would refer to specific applications of those truths. It would refer to the consequences of contingent truths. More precisely, it would refer to truths, e.g. the consequent of (iii), that follow from truths, e.g. the antecedent of (iii) that describe linguistic conventions that happen to hold but needn’t do so. Be it noted that, although the consequent of (iii) follows necessarily from the antecedent, the antecedent is contingent and so is the consequent. It is a contingent fact, as we’ve noted, that (for any appropriate value of P) “if P, then not-not P” is an expression, let alone a true one, let alone a formally true one. And it’s a contingent fact that $E_1 \ldots E_n$, these being the rules constitutive of the English language, are in fact operative. So if logical truth coincided with formal truth, the truths recognized and operated on by those gifted with logical acumen would in all cases be contingent. But logic is the study of *necessary* truth (or, in any case, of a certain kind of necessary truth), and logical acumen is therefore the ability to identify and delineate the consequences of such truths.

4.2 The doubly relative nature of formal truth Consider the sentence:

(a) “Jim is tall iff Jim is tall”

(a) is formally true. (Or so we’ll assume for the time being. In Section xxx, we’ll see reason to qualify this assumption, if not to reject it.) By contrast,

(b) Jim is tall or Jim is tall,

is not formally true. (b) is equivalent with “Jim is tall” and (if true) is therefore empirically true.

Any instance of “P iff P” is formally true since all its instances are true. But not all instances of “P or P” are formally true, since some[361] of its instances are false, e.g. “Jim is a square circle or Jim is a square circle”. (b) is either contingently true or contingently false, since some of its instances are true (e.g. “Paris in France or Paris is in France”) and some are not (e.g. “Paris in Germany or Paris is in Germany”).

But

(AT_1[362]) “all statements of the form *P iff P* are formally true”

is not *itself* a formal truth. After all, it has the same form as:

(TA_1) “all statements of the form *P iff P* are empirically true,”

which is false.

Given any formal truth E belonging to any language L, the proposition *it is formally true that E is formally true* is false. “Either it’s raining here or it’s not raining here” is formally true. But

(IR[363]) “the sentence ‘either it’s raining here or it’s not raining here’ is formally true,”

though true, is not formally true. It has the same form as:

> (RI) “the sentence ‘either it’s raining here or it’s not raining here’ is obviously false,”

which is false.

It is always, in the final analysis, on the basis of knowledge of non-formal truths that we know a given truth to be formal or that we know all statements of a certain form to be true. It’s an empirical fact that “and,” “iff,” “or,” and “because” have the syntactic and semantic properties that they do. But it’s non-empirically true that, *given* that they have those properties, the first three are truth-functional and the fourth is not. That truth is analytic but non-formal. And so is the truth that, the relevant syntactic rules being what they are, “P iff P” is always true, “P or P” and “P and P” are sometimes true and sometimes false, and “P because P” is (I should think) always false.

There is no way to formalize the insights on the basis of which we know that “P iff P” (“P and P”) is always (sometimes). *Given* that knowledge, one can sometimes determine whether a given sentence is true or what it entails. [364] But it couldn’t possibly be through formal operations (operations that are driven strictly by sentence-forms) that one learned which such operations were (or were not) truth-preserving. Given that such truths, in addition to being informal, obviously aren’t empirical, it follows that they are of a non-formal, analytic truths.

Those non-formal truths must themselves be considered part of logic. One of the logician’s jobs is to identify those principles with respect to which arguments are to be evaluated. The formal statements in which the logician’s work may sometimes eventuates don’t constitute such principles. They represent applications of those principles, relativizations of those principles to boundaries set by sets of semantic rules that hold contingently and whose

consequences therefore hold contingently: *given* that a certain semantic rules hold, those principles entail that such and such statements (belonging to the language constituted by those rules) are formal truths. Thus, those formal truths are no more the essence of logic than truths of engineering are the essence of physics. (Engineering: applied physics. Formal logic: applied logic. Not pure logic.)

Since those formal truths are parasitic on, and are to be established on the basis of, other, more fundamental non-empirical truths, it must be in these other truths that the essence of logic is to be found. For---and this much must be said for Frege's conception of logic---logic cannot be parasitic on any other discipline. If we describe a given discipline D as "logic," and then find that D's results are categorically to be interpreted in terms of those of some other discipline D*, then it is D* (assuming that D* itself is not another parasite) that is to be described as logic. So we must regard those logicians who dedicate themselves exclusively to formal work as applied logicians and thus not logicians in the truest sense. And even though not so much as a line of formal reasoning is to be found in Bradley's *Logic* or Mill's *Treatise on Logic*, such authors *were* trying to delineate the principles with respect to which inferences are to be evaluated, and we must therefore regard those authors as being quite as deserving of the title of "logician" as Church and Rosser.

5.0 An alternative analysis "S is formally true"= "S is an instance of an open sentence all of whose instances are true" [365] A sentence is formally true if it is an instance of an open-sentence that is true for all values of its free variables. (An "open-sentence" is an expression that contains a free variable, and is therefore neither true nor false, but is otherwise just like a sentence. So

"*x* is tall" is an open-sentence.[366]) For example:

(5) If Smith is at home, then it's not the case that Smith is not at home;

is a formal truth because it's an instance of the open-sentence;

if *x* is phi, then it's not the case that *x* is not phi,

which is true for all values of *x*. (We'll see that (5) is an instance of many such open-sentences. But its being an instance of just one is enough for its being formally true.) Given any formal truth, the same holds.

If an open-sentence is true for all values of its variables, we'll say that it's "universally true." So "*x* = *x*" is universally true (that being why "Smith = Smith" is "formally" true). It must be kept in mind that this is just an abbreviation. "*x* = *x*" is neither true nor false. No open sentence is true or false.

In light of these points, consider the following expressions:

(1) Bill Gates is wealthy.
(2) *x* is wealthy.
(3) Bill Gates is phi.
(4) *x* is phi.

(1) is a true sentence. (2)–(4) are neither true nor false. Given (1), (2) is what results when "Bill Gates" is replaced with a variable, (3) is what results when "wealthy" is replaced with a variable, and (4) is what results when both are replaced with variables. (See Chapter 7.) Of course, actual sentences are

formed out of these open-sentences by replacing the variables in (2)-(4) with constants. In each case, this procedure may yield either a true sentence or a false one. For example, replacing the variable in (2) with “Bill Gates” yields a truth and replacing it with “Mike Smith,” the name of an impecunious friend of mine, yields a falsehood.

Bearing these points in mind, consider the following expressions:

(5) If Smith is at home, then it’s not the case that Smith is not at home.
(6) If *x* is at home, then it’s not the case that *x* is not at home.
(7) If Smith is phi, then it’s not the case that Smith is not phi.
(8) If *x* is phi, then it’s not the case that *x* is not phi.
(9) If *P*, then it’s not the case that not-*P*.

Each of (6)–(9) is formed by replacing one or more expressions in (5) with a variable. And the following sentences result when, in each case, the variables are bound by universal quantifiers.

(6_{UG}) For any individual *x*, if *x* is at home, then it’s not the case that *x* is not at home.
(7_{UG}) For any property phi, if Smith is phi, then it’s not the case that Smith is not phi.
(8_{UG}) For any individual *x* and any property phi, if *x* is phi, then it’s not the case that *x* is not phi.
(9_{UG}) For any proposition *P*, if *P*, then it’s not the case that not-*P*.

This suggests that a sentence is formally true if it's an instance of a universally true open-sentence.

To take another example:

> (10) If it's necessarily the case that Smith is a mammal if Smith is a dog, then it's necessarily the case that Smith is a mammal if it's necessarily the case that Smith is a dog;

is formally true, because it's an instance of the following universally correct open-sentences:

> (11) If it's necessarily the case that *x* is a mammal if *x* is a dog, then it's necessarily the case that *x* is a mammal if it's necessarily the case that *x* is a dog.

> (12) If it's necessarily the case that Smith is a mammal if *P*, then it's necessarily the case that Smith is a mammal if it's necessarily the case that *P*.

> (13) If it's necessarily the case that *Q* if *P*, then it's necessarily the case that *Q* if it's necessarily the case that *P*.

5.1 What is a logical constant? Quine's analysis, we noted, is vitiated by his failing to say what conditions an expression must satisfy to be a "logical constant." We are in position to provide this information. The expressions "not" and "if . . . then . . ." are on Quine's list of logical constants, and those expressions occur in:

(5) If Smith is at home, then it's not the case that Smith is not at home.

Before we proceed, we have to make a couple of points about the word "not." That word *not* is known as a "sentence-level operator": it's an expression that forms sentences out of sentences—it forms "snow is not white" out of "snow is white." It isn't often that sentence-level operators are replaced with variables. But not only *can* such replacements be made; they must be made. Otherwise, it would be impossible to express any generalization about sentence-level connectives. (Logicians and linguists often make statements about whole classes of sentence-level connectives, and such statements involve replacements of this kind, since they have the form: "given *any* sentence level connective *X* . . .") In fact, we're about to state just such a principle right now.

Given (5), if we replace "not" with a variable, the result is:

(5_{X}) If Smith is at home, then it's *X* the case that Smith is *X* at home.

If we bind the free variable with a universal quantifier of the appropriate kind, the result is:

(5_{XUG}) Given any sentence-level operator *X*, if Smith is at home, then it's *X* the case that Smith is *X* at home,

which is false, the reason being that some sentence-level operators lack the relevant property. An example of such an operator is the word "necessarily."

When this word replaces the variable in (5_X), the result is:

> (5_{XN}) If Smith is at home, then it's necessarily the case that Smith is necessarily at home,

which is false.

Another expression on Quine's list that occurs in (5) is "if . . . then . . ." This expression, like "not," is a sentence-level operator. But whereas "not" is a *one*-placed sentence-connective, "if . . . then . . ." is a *two-place* sentence connective. So, whereas "not" assigns sentences to single sentences, "if . . . then . . ." assigns sentences to ordered pairs of sentences. Given the ordered pair <Toby is a dog, Toby is a mammal>, "if . . . then . . ." yields the sentence: "if Toby is a dog, then Toby is a mammal."

Everything just said about "not" is true of it. If it's replaced with a variable, which is then bound by a universal quantifier of the appropriate kind, the result is:

> (5_{IT}) For any two-place sentence-connective *X* <Smith is at home, it's not the case that Smith is not at home>

But (5_{IT}) is simply false. Let *C* be a two-place sentence-level operator defined as follows: Given two propositions (or sentences) *P* and *Q*, *C*(*P*,*Q*) is true exactly if *P* is incompatible with *Q*. In that case:

> (5_C) *C*<Smith is at home, it's not the case that Smith is not at home>

is false.

What we just said about “not” and “if . . . then . . .” is true of all the other items on Quine’s list of “logical constants.” And it’s true of many other expressions that Quine did not personally include on that list but that, according to many, nonetheless belong on it. (For example, some hold that “necessarily” and “possibly” belong on that list.) This suggests that, when Quine judged an expression to be a “logical constant,” he did so on the (cogent) grounds that replacing occurrences of it with variables systematically failed to yield universally true open-sentences.

This suggests that the following definition of “logical constant” might be the right one (we’ll soon see that it’s only approximately right):

> (LC[367]) For an expression *E* to be a “logical constant” is for there to be some universally-true open sentence that contains *E*.

LC is close to the truth, but it falls just short. Consider the open sentence:

> (#) If *P* and *Q* and grass is green, then grass is green.

(#) is universally true. But, since “grass is green” isn’t a logical constant (#) is false. But, apart from “*P*,” “*Q*,” and “*R*,” all the expressions occurring in:

> (##) If *P* and *Q* and *R*, then *R*.

are logical constants. (##) is universally true, and it contains no expressions that, when replaced with free-variables, result in some *other* open sentence that is universally true. If a universally true open sentence contains no such variables, let us say that it is “pure.” This shows that a “logical constant” is

any expression other than a variable that occurs in a pure, universally true open sentence.

5.3 Context-sensitivity in relation to formal truth Following Strawson (1950), Kaplan (1977, 1989, 1989b) and Barwise and Perry (1983) make the profound point that xxx

Barwise, Jon. 1989. *The Situation in Logic.* Palo Alto, CA: CSLI Publications.

Barwise, Jon, & Perry, John. 1983. *Situations and Attitudes.* Palo Alto, CA: CSLI Publications.

Davidson, Donald. 2004. Problems of Rationality. New York: Oxford.

Fodor, Jerry. 1987. *Psychosemantics.* Cambridge, MA: The MIT Press.

Fodor, Jerry. 1987b. Individualism and Supervenience. In A. Pessin and S. Goldberg (Eds.), *The Twin-Earth* Chronicles, (pp. 192–218). Armonk, NY: M. E. Sharpe.

Fodor, Jerry. 1990. A Theory of Content and Other Essays. Cambridge, MA: The MIT Press.

Fodor, Jerry. 1994. Fodor. In S. Guttenplan (Ed.), A Companion to the Philosophy of Mind, (pp. 292–300). Oxford: Blackwell.

Frege, Gottlob. 1884. Die Grundlagen der Arithmetik: eine logisch-mathematische Untersuchung über den Begriff der Zahl. Breslau: W. Koebner. Translation: J.L Austin 1974. The Foundations of Arithmetic: A logico-mathematical enquiry into the concept of number, 2nd ed. Blackwell

Gödel, Kurt. 1953. Is mathematics syntax of language? In S. Feferman (Ed.), Kurt Gödel: Collected Works, Volume III, (pp. 334-356). New York: Oxford.

Hempel, Carl Gustav. 1965. *Aspects of Scientific Explanation.* New York: Free Press.

Hempel, Carl Gustav. 1966. *The Philosophy of Natural Science.* Englewood Cliffs, NJ: Dover.

Kaplan, David. 1989a. Demonstratives. In J lmog, et al. (Eds.), Themes from Kaplan (pp. 481-563). New York: Oxford University Press.

Kaplan, David. 1989b. Afterthoughts. In J. Almog, et al. (Eds.), Themes *from Kaplan* (pp. 565-614). New York: Oxford University Press.

Pap, Arthur. 1958. Semantics and Necessary Truth. New Haven: Yale University Press.

Pylyshin, Zenon. 1984. Computation and Cognition: Toward a Foundation for Cognitive Science. Cambridge MA: MIT Press.

Quine, Willard van Orman. 1940. *Mathematical Logic*. Cambridge, MA: Harvard University Press.

Quine, Willard van Orman. 1966. *The Ways of Paradox.* Cambridge, MA: Harvard University Press.

Quine, Willard van Orman. 1970. *Philosophy of Logic.* Cambridge, MA: Harvard University Press.

Russell, Bertrand. 1903. *Principles of Mathematics.* Cambridge: Cambridge University Press.

Russell, Bertrand. 1920. Introduction to Mathematical Philosophy. London: George Allen Unwin.

Strawson, Peter. 1950. On Referring. In A.P. Martinich (Ed.), *The Philosophy of Language,* (pp. 219–234). New York: Oxford University Press.

Strawson, Peter. 1952. An Introduction to Logical Theory. London: Methuen and Co.

Wittgenstein, Ludwig. 1922. *Tractatus Logico-philosophic us.* London: Routledge, Kegan & Paul.

Wittgenstein, Ludwig. 1958. *The Philosophical Investigations.* Oxford: Blackwell.

Taylor, Kenneth. 1998. *Truth & Meaning.* Oxford: Blackwell.

Time-travel: An incoherent concept

Setting aside the trivial and irrelevant fact that we travel forward in time at a rate of one minute/minute, the statement "time-travel is possible" has three distinct meanings:

> (a) One can "leap" from one time to some non-adjacent time; i.e., one can move in a *discontinuous* manner from one time to another.
> (b) One can travel forward in time at a rate of *more* than one minute/minute.
> (c) One can travel *backwards* in time.

It is generally thought that (a)–(c) are empirical questions that fall within the jurisdiction of physics. In actuality, they are questions of logic, not of empirical fact. Moreover, each of (a)–(c) is logically false and therefore couldn't possibly be correct.[368]

Let's discuss (a) first. In this chapter, we have seen that anything that persists for any amount of time is a causal series**.** A causal series is a continuous series of states of affairs such that, given any two of them, one is the (in)direct cause of the other.

An illustration may be helpful. At any given moment during its existence, your car consists of an incessant flurry of mass-energy displacements; and your car is thus a series of "event-flurries," as we'll call them. Twenty-four hours ago, your car was constituted by one event-flurry. Right now it's constituted by a different event-flurry. (Let S_1 be the first event-flurry and let S_2 be the second.) And, of course, my car is currently constituted by some

event-flurry. (Let S_3 be that flurry.)

Bearing all of this in mind, imagine the following. While your car was parked in your driveway, somebody crashed into it, severely damaging it. Your car therefore bears little resemblance to what it was twenty-four hours ago. It follows that S_1 and S_2 are qualitatively very different from each other. By contrast, S_1 and S_3 are as similar to each other as their being in different places permits. This is sheer coincidence: S_1 and S_3 are, we will suppose, as causally disconnected from each other as any two non-simultaneous events on this planet can be. It follows that my car, as it is right now, is just like your car, as it was twenty-four hours ago, before it was damaged. It also follows that S_1 is much less similar to S_2 than it is to S_3.

Be all of this as it may, the car in your driveway is, whereas the car in my driveway is not, numerically identical with the car you owned twenty-four hours ago. Mediating between S_1 and S_2 is a continuous series of events such that, given any two of them, one is the (in)direct cause of the other. No such series mediates between S_1 and S_3. That's why (i) the car in my driveway isn't the very same car as the one you parked in your driveway twenty-four hours ago, their uncanny mutual resemblance notwithstanding. And that's also why (ii) the car sitting in your driveway is the very same car as the one you parked in your driveway yesterday, their *lack* of mutual resemblance notwithstanding. Obvious generalizations of these points establish our thesis that objects are identical with causal series.

A consequence of the fact that objects are such series is that a given object ceases to exist the moment the corresponding causal series is severed. If, instead of just being damaged last night, your car was vaporized, my car, though an exact duplicate of your (pre-vaporized) car, still isn't numerically identical with it.

Given that an object ceases to exist the moment the corresponding causal

series is severed, it follows that, so far as it isn't logically false, the statement:

> (i) Smith disappears in 2009 and re-appears in the year 2023.

collapses into the statement:

> (ii) Smith is annihilated in 2009 and a Smith-*like* entity that isn't actually Smith *ex nihilo* comes into existence in the year 2023.

A question will help clarify the reasoning behind this last assertion: Where is Smith during that fourteen year interval? He's either nowhere or he's somewhere. Let's suppose that he's nowhere. In that case, for the reasons illustrated in connection with your car, whatever it is that pops into existence in the year 2023 isn't numerically identical with Smith. So let's suppose that he's somewhere—that, at any given instant during that time period, he is in some place or other. But in that case he's no more time-traveling than anything else that exists during that interval. If he's next to some one tree during that entire period, he's no more time-traveling than that tree. If he's going from place to place, he's no more time-traveling than any one of the satellites that will orbit the earth during that period.[369]

So depending on whether he's somewhere or nowhere, Smith is either permanently annihilated in 2009 or he's no more time-traveling than anything else that exists during that period. Therefore, (i) is false.

To see why (ii) is false we need do little more than repeat some of the points just made. Suppose that Smith travels from 2009 to 2023 at a rate of one million minutes/minute. We've just seen that Smith cannot make this journey in a discontinuous manner; he cannot disappear from one time and

reappear at some other, non-adjacent time. So Smith must make this journey in a continuous manner. So at any given moment during this interval, Smith must be in some place or other. It follows, as we saw a moment ago, that he is no more time-traveling than anything else that exists during that interval. The supposition that Smith is traveling forward in time at a rate of one million minutes/minute thus turns out to collapse into the supposition that he isn't time-traveling at all. Given any positive number *n*, an analogous argument shows that continuous forward time-travel at a rate of *n* minutes/minute collapses into forward time-travel at a rate of one minute/minute.

There is another, quite different problem with the concept of discontinuous time-travel. Let W_1 be a universe where (i) is true and (ii) is false; and let W_2 be an otherwise indistinguishable universe where (ii) is true and (i) is false. What happens in the one universe at any given time coincides with what happens in the other universe at that time. It follows that, even if (*per impossibile*) an occupant of the one universe knew exactly what happened in that universe at any given time, he would not, in virtue of that fact, have any legitimate reason for preferring (i) to (ii). *A fortiori*, no possible observation that one might have would provide one with any reason for believing oneself to be in W_1 as opposed to W_2. Since observation couldn't possibly provide one with such a reason, only logic could do so. But logic cannot do so, since (i) is logically false, as we've already seen. [370]

Similar points hold in connection with discontinuous backwards time-travel. In the movie *Back to the Future*, young Marty McFly vanishes from the year 1985 and shows up in the year 1955. This means that, in the year 1985, the causal sequence that constitutes his existence is severed. Like everything else that persists for any length of time, Marty McFly is a causal sequence. It follows that, so far as it isn't incoherent, the statement:

> (1) In 1985, Marty McFly traveled backwards in time to the year 1955 and is otherwise unchanged.

collapses into:

> (2) Marty McFly was annihilated in 1985, and in 1955 there appears *ex nihilo* somebody who is numerically distinct from him but otherwise just like him.

Also, given obvious adaptations of what we said a moment ago, there is no conceivable event that, were it known, would provide more support for (1) than for (2).

Since, therefore, there couldn't possibly be any observation-based reason for preferring (1) to (2), the only possible reasons for doing so would be of a logical nature. This means that there are *no* reasons for doing so, given that (1) is incoherent. The concept of discontinuous backwards time-travel is therefore a broken one.

What about the idea of *continuous* backwards time-travel? For argument's sake, let's suppose that, during his continuous journey from the year 1985 to the year 1955, Marty McFly is right next to some stop-sign that was built in 1920 and stays in the same place for several centuries. That means that, at any given moment during that thirty year period, Marty McFly is no more time-traveling than the stop-sign. Each is just sitting there.

There are, of course, some important differences between Mary McFly and the stop-sign. First of all, whereas the stop-sign is crushed or burnt or otherwise destroyed in some pedestrian way, Marty McFly must simply vanish in the year 1985. (If he doesn't, then he's still hanging around in the year 1985, in which

case he obviously hasn't gone back in time.) And, whereas the stop-sign's presence in the year 1955 is to be explained in the customary way, Marty McFly just popped up *ex nihilo.*

Bearing these points in mind, let W_3 be a universe in which:

> (3) Marty McFly continuously traveled backwards in time from the year 1985 to the year 1955 but is otherwise unchanged.

is true, and in which (2) is false; and let W_4 be a universe that is exactly like W_3 *except* that, in it, (2) is true and (3) is false. Echoing what we said earlier, there is no conceivable event that, were it known, would enable one to determine whether one was in W_3 as opposed to W_4, or *vice versa.* Which, taken together with the fact that being annihilated in 1985 didn't exactly tighten young Marty's grip on existence, suggests that, other things being equal, (2) is the better way to model the observational data.

There is another, even deeper issue. For there to be backwards time-travel of any kind, there must be retroactive causation: some events must be the effects of *later* events. Marty McFly's appearing in 1955 must be an effect of some event occurring in 1985. But there are empirical and also strictly logical reasons for thinking backwards time-travel an impossibility.

If Relativity Theory (RT) is right, which we'll henceforth assume that it is, the statement "E_2 is an effect of E_1, but E_1 post-dates E_2" is tautologically false. According to RT, order in time is *defined* in terms of causation: anything caused by a given event *ipso facto* follows that event. Thus, the view that one could travel backwards in time is in the same category as the view that triangles could have five sides.

A well-known consequence of RT is that a given event may precede one event relative to one framework, be simultaneous with it relative to some

second framework, and follow it relative to some third framework. Some conclude from this that backwards time-travel *is* possible.

This is not the right inference to draw. For reasons that aren't important here, the framework-relative nature of temporal order is a *consequence* of, among other things, the fact that any effect of an event *ipso facto* follows that event. In other words, it's a consequence of the fact that "E_2 precedes E_1 and is also an effect of it" is tautologically false.

Like any other scientific theory, RT is empirical. It's based on observation; and acceptance of it is warranted only as long as our observations are consistent with it. Nonetheless, the principle that causes ipso facto precede their effects is an analytic, not an empirical, truth. (We'll henceforth refer to this principle "CT"—short for "causality and temporal order.") So CT is true even if RT is false.

"But how can CT be analytic," it will be asked, "given that RT is an observation-based theory?" Two reasons. (1) Not every statement composing an empirical (observation-based) theory is *itself* observation-based. Every theory comprises statements of pure logic (e.g., "no true statement entails the negation of anything entailed by any of its consequences") that aren't observation-based. (2) Not everything learned with the help of observation is learned *through* observation. If one learns some complex mathematical theorem by reading a proof of it, one learns that theorem *through* observations of ink-deposits; but it isn't *by observing* such deposits that one learns it. The observations of ink-deposits merely occasion or "trigger" the purely logical ratiocinative activity that *does* lead to a genuine knowledge of that principle.[371] Somebody with 20/20 vision who has carefully studied those ink-deposits (i.e., who has studied those ink-deposits themselves, as opposed to their meanings) may completely and utterly fail to understand the corresponding proof; whereas somebody whose vision is just good enough to

see some of the key lines of the proof may, on the basis only of those observations along with the thoughts they precipitated, understand it extremely well.

Also, somebody who has heard the proof stated but hasn't read it may, on the basis only of those auditory observations along with the thoughts they precipitated, thoroughly understand the proof. Ink-deposits bear no physical resemblance to noises. Therefore, it isn't by studying ink-deposits that one learns the proof.[372] So given only that empirical discoveries led to our becoming aware of CT, it doesn't follow that CT was itself discovered empirically.

And, in fact, CT couldn't have been discovered empirically. Assuming it a given that it was Smith's pushing the button that caused the elevator to come, there is no conceivable event that, were it known, could warrant a revision of one's subsequent judgment that the pushing of the button preceded the arrival of the elevator.

Einstein himself very clearly said that CT *was* not an empirical hypothesis. He also said (unfortunately for me) that it was a convention—a *definition* of "order in time"—and, therefore, not a substantive truth. I would suggest that, although he described CT as a definition, he chose this particular definition, as opposed to others, because he saw in it a substantive truth whose roots lie in the very concept of what it is for one event to precede another.

Davies, Paul 1996. About Time. Pocket Books.
Davies, Paul 2002. How to Build a Time Machine. Penguin Books Ltd.
Einstein, Albert. 1962. The Principles of Relativity. New York: Dover.
Einstein, Albert, & Infeld, Leopold. 1961. The Evolution of Physics. New York: Simon & Schuster.

Fodor, Jerry. 1981. Representations. Cambridge, MA: The MIT Press.
Gale, Richard M. 1968. The Philosophy of Time. Palgrave Macmillan.
Gott, J. Richard. 2002. Time Travel in Einstein's Universe: The Physical Possibilities of Travel Through Time. Boston: Mariner Books.
Gribbin, John. 1985. In Search of Schrödinger's Cat. Corgi Adult.
Lewis, David. 1986. Collected Papers, Vol. II. New York: Oxford.
Miller, Kristie (2005). "Time travel and the open future". Disputatio 1 (19): 223–232.
Nahin, Paul J. 2001. Time Machines: Time Travel in Physics, Metaphysics, and Science Fiction. Springer-Verlag New York Inc.
Nahin, Paul J. 1997. Time Travel: A writer's guide to the real science of plausible time travel. Writer's Digest Books. Cincinnati, Ohio. .
Pagels, Heinz 1985. Perfect Symmetry, the Search for the Beginning of Time. Simon & Schuster.
Pickover, Clifford. 1999. Time: A Traveler's Guide. Oxford University Press Inc, USA.
Poincaré, Henri. 1913. Last Essays. New York: Dover.
Randles, Jenny. 2005. Breaking the Time Barrier. Simon & Schuster Ltd.
Reichenbach, Hans. 1956. The direction of time. University of California Press.
Reichenbach: 1958, The philosophy of space and time. Dover.
Russell, Bertrand. 1928. The Analysis of Matter. London: George Allen and Unwin.
Russell, Bertrand. 1948. Human Knowledge: Its Scope and Limits. London: George Allen Unwin.
Shore, Graham M. "Constructing Time Machines". Int. J. Mod. Phys. A, Theoretical.
Toomey, David. 2007. The New Time Travelers: A Journey to the Frontiers

of Physics. New York: W.W. Norton & Company.

Probabilistic Causation

Note: What follows is my attempt to state, in as clear a maner as possible, the contents of a lecture given By Paul Humphries. I deserve partial credit for the contents of this paper, in as much as I made (what I believed to be) the needed amendments and additions to Humphries' paper. But Humphries' own work must be examined before a judgment can be made as to who deserves credit for what.

According to Leibniz, the two cornerstones of cogent thought are the law of non-contradiction (not both P and not-P) and the "principle of sufficient reason" (psr). psr is ambiguous, it can mean either:

(i) any given event has a cause, meaning that, if E2 is an event, there is a prior event E1 such that, given E1, E2 *had* to happen;

or

(ii) there is a *justification* (a *reason*) for the occurrence of any given event.

Be it noted that it is events (or conditions) that have causes, whereas it is propositions that have justifications.

Whether (i) is true is not the concern of this paper. Until I recently heard a lecture by Paul Humphries, I accepted (ii). I now reject, on the basis (partly) of what Humphries said. [373] What I wll say, though in some m ways a development of what Humphreys said, is not Humphries-exegesis, incorporating into it large elements of my own reactions to his talk.

Humphreys' denial of (ii) consisted in his advocating the contention that:

(H) a given event E1 can cause (or the affirmation of its existence can justify) the occurrence (or affirmation thereof) of an event E2, even if, given E1's occurrence, there was only an n% chance (where n is any number greater than 1 and less than 100) of E2's occurrence.

I believe H correct and would like to defend it, borrowing heavily (to say the least) in my attempt to do so from Humphries' argument.

A preliminary argument. Complying with laws is different from being governed by them. Everything complies with the laws of microphysics; and, presumably, all higher level laws (assuming such thing to exist, which I will argue they do) they hold in virtue of microphysical laws.

But consider this. We have two houses that are architecturally identical, meaning that they way in which their medium sized parts (their constituent bricks, planks, etc.) are interrelated are identical. Despite that fact, the microphysical structure of any component of either house is completely different from the corresponding structure of any component of the other house. However, these differences have no bearing on the dynamic interrelations on the bricks, planks, etc. which those microrealities constitute.

Both houses fall. They both fall for the same reason. That reason has nothing to do with the microstructures of their parts. It has to do with the fact that some plank wasn't properly nailed to some other plank, or some such.

The mechanical (Newtonian, macro-level) laws involved in both cases were the same. At the same time, in both cases, those laws were operative only because of facts about the relevant microconstituents, which, of course, obey a completely different set of laws.

We thus have a choice. On the one hand, we can say:

(a) The laws compelling the collapse of both houses differed. The reason: the microconstituents, and the dynamic interrelations holding among them, differed; and, at the end of the day, everything supervenes on micro-reality.

On the other hand, we can say:

(b) In both cases, the same laws were involved; but those laws were *implemented* by different microphysical mechanisms.

To my mind, (b) is obviously the right option. If we go with (a), we make it a mere accident or coincidence that both houses fell, even though it was clearly no accident.

This brings us to the next fact. Most laws hold only probabilistically. It's a theoretical possibility that two situations that, from a Newtonian perspective are indistinguishable, could (*ceteris paribus*) have different outcomes---because of some quantum event that happened to ramify outside the normal sphere of influence of such events. (It is possible, given quantum mechanics, and also consistent with Newton's laws, so far as the latter are consistent with quantum mechanics, that I could walk through a brick wall.)

Given this, we *could* say:

(c) Since, given initial conditions C1, Newtonian laws don't provide a 100% guarantee that non-iniital conditions C2 will obtain, we can't *really* say C1 brought about C2. So far as we did say such a thing, it would be shorthand for some unwieldy, but correct, microphysical statement.

But if we went with (c), we'd make it impossible to produce precisely the sort of explanatory meaningful generalization discussed earlier in connection with the two houses. Which means that we'd make it impossible to produce otherwise easily identified and illuminating explanations.

Here we have a choice. We can say either

(#) The structural (non-microphysical) defects in the houses *caused* them to collapse'

or

(##) Those defects *did* not cause those houses to collapse, since causes operate with 100% efficacy, and those defects don't *always* lead to outcomes such as those described.

If we go with (#), we satisfy the explanatory desiderata described earlier. If we go with (##), we hold onto our *a priori* belief that, if E2 results from E1, then, given E1, E2 couldn't have failed to occur. This position, though not to be jettisoned lightly, is so inimical to explanation that it must, in fact, be jettisoned.

Newton's laws, even after recent development in physics are taken into account, hold almost 100% of the time. So let's consider some laws (or 'principles,' if you consider my use of the word 'law' question-begging) that are very obviously probabilistic.

Smoking causes cancer. This would seem to be a law, or principle, of sorts. But less than 30% of smokers get lung cancer. Nonetheless, if asked, "why did Jerry get lung cancer?,"

We are (let's assume) correctly answering the question by saying "because he smoked 4 packs a day." So here we can say that

(E1) Jerry's heavy smoking

caused

(E2) Jerry's cancer,

even though events relevantly like E1 have only a small chance of bringing about evnetws relevantly like E2.

Let us now suppose that Larry smokes as much as Jerry but doesn't get cancer. In that case, it's hard to see how E1's being the cause of E2 could be maintained. "What we must say," it will be said, "is that E1 did not *by itself* cause E2---that there was some other factor."

I agree. There *was* surely some other factor. (To facilitate discussion, let's suppose that other factor to be one that falls totally outside the reach of medicine—that it was some quantum-event that over-ramified.)

But biology (medicine) is an attempt to provide explanations. As such, it attempts to produce propositions that bear on one another in ways that, given data (falling into certain categories), light can be shed on that data's existence. The law (if I may so speak) that smoking causes lung cancer is not an ideal law; too many hidden variables are unaccounted for. But there would (let us suppose, going on the hypothesis that those hidden variables belong to quantum physics, not biology) be no way to include those hidden variables in biologically meaningful generalizations. Let's say that, in Jerry's case, there occurred some quantum event Q, the likes of which did not occur in Larry's case, and that, when combined with Jerry's habits, Q led to his getting cancer.

When explaining why *Jerry* got cancer, we can mention Q, and in doing so we'll be providing a much better explanation than we'd otherwise be.

But neither Q, nor the likes thereof, has any place in any *biological generalization.* So, when explaining why *Fred*'s smoking led to his lung cancer, the fact that Q played a role in Jerry's case is explanatorily idle. So, though causally operative in Jerry's case, there is a sense---an epistemic, not an ontological sense---in which Q was irrelevant. It was irrelevant in that, *qua* situation capable of being biologically explained, Q did nothing.

That said, it would be absurdly revisionist to say that Jerry's smoking habits didn't cause him to get cancer. Since such habits lead to lung cancer only a fraction of the time, we must say that there can be such a thing as probabilistic causation, meaning that E1 can be the cause of E2,even though, given E1, there was a less than 100% of E2's occurring.

Humphreys underemphasized an important point. For the reasons just given, there is no doubt that *epistemically speaking*, i.e. speaking from the viewpoint of explanation, there is such a thing as probabilistic causation. But that's a very different thing from saying that there is such a thing that, *ontologically* speaking, there is such a thing as probabilistic causation. The orthodoxy these days is that, ontologically, there is such a thing as probabilistic causation. (Incidentally, John von Neumann's proof that the introduction of hidden variables could not make quantum physics be deterministic was shown, by David Bohm (1958), to establish that quantum physics could not be made *epistemically* deterministic, it being left open by von Neumann's proof whether it could still ontologically deterministic.)

To move on, I must make some foundational points. Scientific explanations justify events. This means, among other things, that they to be capable of integrated, in an explanatory sense, into comprehensive regularities or *laws.* [374]

Justification is a relationship that holds between propositions, not between facts. By a “fact” (or “situation”) I mean actual assemblages of objects, e.g. the chair’s being next to the cat, which is on top of the carpet, etc. In other words, I mean a (re)distribution of mass-energy *in virtue of which* a given proposition holds, but which is not itself true or false. Facts exist or they don’t; they’re not true, even when existent.[375]

Facts don’t *bear* on anything. (That is, they don’t bear on other things in a logical or confirmational sense; obviously they bear on them in a causal sense.)

Propositions consist of concepts. (By “concepts” I mean properties. So I mean concepts in the objective, Fregean sense—cf. “anything falling under the concept *whale* falls under the concept *mammal*,” and I therefore don’t mean them in the subjective, Kantian sense---cf. “Smith’s concept of space-time is less developed than Einstein’s”).[376] Facts consist of object (instances of concepts). A (non-analytic) proposition is a conceptualization (‘propositionalization’) of a given fact. A given fact admits of indefinitely many propositionalizations. I walk across the room. The mass-energy displacements constitutive of that facta are necessary and sufficient for many different kinds of propositions--- atomic, chemical, physiological , psychological.

Two propositions, P1 and P2, can hold in virtue of the very same fact, and yet be such that neither has the slightest explanatory bearing on the other. It may be that for each psychological truth that holds in virtue of my deliberately walking across the room there is a corresponding biochemical truth. But *chemical reaction xyz occurred* isn’t in any obvious way explanatorily relevant to *JMK felt it in his professional interest to stampede in a visibly angry fashion across the room.*

Scientific disciplines presumably aspire to identify truths. But their *first* objective must be to model data. It's only through data, after all, that truths can be known. (And there must be some way, *without* knowing the relevant truths, of evaluating the viability of a given model, since it is only through such truths that those truths are known.)

Disciplines consist of propositions, not of facts. Those propositions, when true, hold in virtue of facts. But they are not themselves facts. The proposition *the gray rock has a larger mass than the blue* is different from the *fact* that the *ceteris paribus* more energy is needed to displace the first rock by a given amount than the second. As previously stated, propositions consist of concepts, not of (spatiotemporal) objects. A scientific discipline is an attempt to model a certain body of data. Data can be thought of as propositions (truths[377]).

Disciplines are individuated primarily by the *data* that they aspire to model, and only secondarily, if at all, by the truths underlying those data.[378] Dogs and plants are physical objects and no more fall outside the scope of physics than rocks. But biology and botany are not physics. How is this possible? Biological data is not physical data. (By 'physical,' I mean 'relating to the discipline of physics, not 'spatiotemporal' or 'extended' or anything of the sort.) Strictly speaking, it isn't' dogs whose behavior falls within the scope of physical laws. It is the heaps of particles that *constitute* dogs that do so. Such heaps aren't dogs, since the heap can undergo qualitatively changes (redistributions of its constituents) without the dog's undergoing any (biologically meaning, dog-relevant) changes. Thus, *qua* dog-constituting heaps, those heaps fall within the scope of biology, not of physics. *Qua* mere (non-dog-constituting) heaps, they don't fall within the scope of biology, belonging exclusively to physics.

Disciplines model data (propositions). To model propositions is to systematize them in such a way that their bearing relations on one another are identified and appropriately organized. Propositions about facts are exceedingly crude representations of those facts. Consider the proposition:

(*) JM walked across the room.

(*), though true, tells us nothing about the myriad atomic, chemical, and physiological events in virtue of which (*) is true. Propositions consist of tendentious conceptualizations of bodies of data: of inquiry-driven over-consideration of certain aspects of those data and under-consideration of other aspects of them. So I is inherent in what propositions are that they leave out facts. Those facts are often causally relevant. But they cannot always fit into the system propositions constitutive of a given discipline. The generalizations constitutive of disciplines are constructed of concepts under which those facts may not fall. Therefore, the law (or principles, if you prefer) constituting those disciplines may not acknowledge the existence of those causally efficacious factors. So, explanatorily speaking, those factors may not exist, even though, ontologically speaking, they do exist.

Science, so it is often said, has no interest in what is not operational. Since a particle's momentum and position cannot simultaneously be determined, the concept of simultaneous momentum and position is without scientific significance. (I do not myself accept or reject this view; it falls outside the scope of this paper.[379]) Assuming this principle correct, then events that, ontologically speaking, are causally efficacious are, explanatorily speaking, non-existent; a consequence of which, as we've seen, is that, at least from a discipline-internal (explanatory) standpoint, such causes don't exist; which, in its turn, establishes Humphrey's claim that (epistemically

speaking) there is probabilistic causation and that, as Humphrey's said at one point, "the principle of sufficient reason" is false. (More precisely, it establishes, or suggests, that its false on its epistemic, not its ontological, disambiguation: a distinction that did not play a big part in Humphrey's paper.)

-

Propositions as Sets of Properties, Truth as Instantiatedness

Consider the proposition

(*) *there is a cactus in front of my window*

A world where (*) is true is different from one where (*) is false. In the one kind of world, there is an instance of the property of being a cactus in a certain place, at a certain time. In the other kind of world, there is no instance of that property. It thus seems that for (*) to be true is for various properties to be instantiated and that for (*) to be false is for those same properties *not* to be instantiated.

As it happens, (*) is false. But I can easily enough imagine what things would be like if it were true. I imagine a world where, at certain places and times, there are instances of certain properties.

What is true of (*) is obviously true of other propositions. These reflections suggest a certain view as to what propositions are, and as to what it is for a proposition to be true. A proposition is a collection of properties, and a proposition is true if the relevant properties are instantiated.

I thus propose that a proposition is a set of properties, and that for a proposition to be *true* is for all of those properties to be instantiated. Truth is identical with instantiatedness.

In this paper, I will develop the suggestions just put forth. There are atomic propositions and molecular propositions. The position in question can be made to apply to both.

§ There are two requirements that any adequate theory of propositions must

satisfy. First, it must accommodate the fact that a propositions has a certain *composition* – or, as Katz put it, a certain *de*composition. Second – this is related – it must account for the *unity of the proposition*; it must explain how a proposition's various constituents can combine into something true or false, instead of remaining a truth-valueless set or heap.

The proposition *Socrates punched Plato* has a unique decomposition into minimal parts; and that proposition is obviously identical with a certain arrangement of those parts. The question is: *how* exactly are those parts arranged in that proposition? What conceivable arrangement of Plato, Socrates, and the relation of punching constitute a true proposition? The set (Plato, Socrates, the relation of punching) obviously isn't true or false. It doesn't matter whether we impose some kind of ordinal structure on that set. The *ordered* triple <Plato, the relation of punching, Socrates> isn't true or false. Neither is <the relation of punching <Plato, Socrates>>, or any other ordering of those three constituents. So even though it seems clear that *Plato punched Socrates* is identical with *some* arrangement of those three things, it is far from clear what kind of arrangement is the appropriate one. In my view, it becomes quite clear what the right kind of arrangement is, so long as we pursue the suggestions made earlier regarding the nature of propositions.

(A) "Smith punches Jones"

is a sentence.

(B) "Smith, the relation of punching, Jones"

is not.

A sentence is not just a heap of referring terms. A corresponding point holds

about propositions.

(Ap) *Smith punches Jones*

is a proposition.

(Bp) *Smith, the relation of punching, Jones*

is not. A proposition is not a heap of objects and properties. A proposition is a set of objects and properties that are *unified* in some way. But *how* must they be unified? In what does the "unity of the proposition" lie? This question has never been adequately answered. I believe that our analysis of propositions gives us an answer to this question.

§ Frege's answer was to say that, in a proposition, certain items are "unsaturated": they have holes in them. These holes enable the items in the proposition to link together in the right way. In (A), the expression "punches" corresponds to something with holes in it; these holes enable Smith and Jones to link up with that thing in the appropriate way. By contrast, in (B), the thing corresponding to "the relation of punching" doesn't have the right number of holes in it; so it doesn't appropriately link up with Bob and Smith.

Frege's "solution" is really just a way of labeling the problem, as Davidson said.[380] Obviously we aren't talking about physical holes. We are talking about holes in a metaphorical sense. What we want to know – what Frege doesn't tell us – is what this metaphor is a metaphor for.

The difference between (A) and (B) obviously has something to do with the fact that the first has grammar, whereas the second does not. In the first, there are inflections. (I will argue later some of these are not phonetically apparent.)

In the second, nothing is inflected. The difference between a proposition and a non-propositional heap is somehow coded in these grammatical articulations. The question is: what exactly these articulations do? Do they themselves refer to entities of some kind? Does the "-s" on "punches" refer to some mysterious entity – perhaps to kind of function? Is *that* why (A) encodes a proposition while (B) does not? Are grammatical articulations *referring terms* with the special property that the things to which they refer confer unity on what would otherwise an unordered heap? Or do grammatical articulations not refer to anything? Do they exert their unifying influence in some other way?

§ Let me start out by stating, in extremely rough outline, the position I will defend. Consider the ordered-set <2,3>. The ordinal properties of this set can be coded in, and actually identified with, the properties of some *unordered* set, namely ((2), (2,3)). I will argue that propositions are unordered sets of properties. Thus *Smith punches Jones* is a set of properties. For a proposition to be *true* is specifically for all of its constituents to be *instantiated.* Truth is instantiation. For a proposition to be true is for all the members of a certain kind of set to be instantiated.

Here two worries immediately arise. First, it seems obvious that *Smith punches Jones* and *Jones punches Smith* have the same *constituents,* even though they are obviously different propositions. I will argue that those two propositions *share* constituents – they have in common Jones, Smith, and the relation of punching. But the one proposition has constituents not had by the others; and the ordinal differences between those two propositions are reflected in these differences in their constituents.

The other worry is this: Surely Smith is not a *property*; surely Smith cannot be instantiated. Let me outline what I will say about this. On my analysis, (Ap) is identical with some set S that comprises only things that can be instantiated,

and that therefore comprises only *properties*. I will argue that what is a member of S is not Smith *per se* but is rather the property of being identical with Smith. Obviously *that* thing can be instantiated.

No doubt, this maneuver will raise another concern: a concern I will spend a lot of time trying to allay. It will at once be said that what is a constituent of (Ap) is *Smith* and not the property of being identical with Smith. (Ap) says that *Smith* punched Jones, not that the property of being identical with Smith did so.

I fully agree: what is a constituent of (Ap) is Smith, and Smith – nothing else – is what punched Jones. But we need to be very clear what it means to say that something is a *constituent* of a proposition. As I will argue, when you say that Smith is a "part" or "constituent" of a proposition P, you are not saying that Smith's relation to P is anything like a brick's relation to a house or a wheel's relation to a car. What you are saying is that the truth of P depends in some way (that we will specify) on Smith: the "constituency" relation in question is a relation of *dependence* – it is a purely logical relation, and cannot be modeled on a spatial relation of any kind. Now suppose we identify (Ap) with some set S, one of whose members is the property of being identical with Smith. By doing so, I will argue, we can do justice to the relation that holds between (Ap) and Smith – the relation of dependency that holds between them. So far as it is meaningful, let alone true, to say that Smith is a "part" or "constituent" of a proposition, what is being said is precisely that there exists just the kind of dependency relation we've been discussing. Our construal of propositions as sets of properties, along with our construal of truth as the instantiatedness of the members of such sets, captures that dependency relation to a tee.

There is more to say on this matter. Consider the following two statements:

(i) x is a constituent of a proposition P.
(ii) y is a member of set S.

The sense in which x is a part of P is entirely different from the sense in which y is part of S. x is a part of P only in the sense that the truth of depends on some fact about x. y is a part of S in some other sense. It is true that I identify

(Ap) *Smith punches Jones*

with some set S; it is also true that the property *identical with Smith* is a member of S, and not Smith himself. But the members of S are not the constituents of P. For Smith to be a constituent of (Ap) is specifically for (Ap) to be about Smith -- it is for there to be a dependence relation between Smith and (Ap). Precisely that dependence-relation holds between Smith and (Ap) if we identify (Ap) with S, and then go on the truth of (Ap) with the being-instantiated of all the members of S. So our analysis *is* consistent with the truth of statement that Smith is "constituent" or "part" (Ap), at least on any significant disambiguation of that statement. It is irrelevant that Smith himself is not a member of S. When we are talking about the membership of S we are talking about one thing; when are talking about the constituency of P we are talking about another (we are taking about relations of dependency or aboutness). There is a *connection* between these two sets of facts: facts about the constituency of P – facts about the aboutness relations in question – are *realized by* or *supervene on* facts about the membership of S. But, strictly speaking, the two sets of facts are distinct. We will explore this further in due course.

§ There is yet another worry. Some may object to the idea that there is such a thing as the property of being identical with Smith. It will be said that

tallness is a property, and so is wetness – but there is no such thing as the *property* of being identical with Smith.

I think this point of view is wrong for two different reasons. First, statements like "so and so is not identical with Smith" are obviously meaningful – indeed, they are often true -- and it is hard to see how this could be the case unless there were such a thing as being identical with Smith.

Second, suppose that Smith exists in some world w. Obviously Smith is not a *basic* constituent of w; he is not one of its lower level constituents – he is not a quark or a muon or a single displacement of mass-energy. All of Smith's activities – indeed, his very existence – supervene on lower-level states of affairs. So *if* Smith occupies some part of space-time, that supervenes on much more basic facts – facts about the displacements of mass-energy (or whatever the basic constituents of the world in question are). Nobody is going to deny the existence, let alone the scientific respectability, of the properties involved in those lower level states of affairs – charge, spin, mass, and so on. Nobody is going to say that there is anything suspect or "viciously circular" about believing in these properties. Every fact about Smith, even his very existence, is fixed by facts about those wholly acceptable, lower-level properties. So the property of comprising a Smith is a perfectly respectable property of a world. It is a necessary consequence of there being certain patterns of mass-energy displacement in that world. Given the respectability of these lower-level states of affairs, and given that Smith's existence is a necessary consequence of them, it is hard to see the property of being Smith in any way lacks respectability – it is hard to see why we should deny existence to the property of being identical with Smith.

We might even think of it this way. It is clear that aesthetic and moral properties are realized by lower-level states of affairs: certain patterns of mss-energy displacements realize beautiful works of art, while others don't. Given

that Smith's existence is realized by lower-level states of affairs -- by certain patterns of mass-energy – we might even think of Smith as a kind of *property.* If a pattern of mass-energy displacements has the right properties – if, perhaps, it has the right origins, and it has the right internal properties – then it "Smiths".

Of course, there are profound differences between the property of being beautiful and the property of Smithing. A world can comprise many beautiful things; but – leaving aside things like cloning, meiosis, and so on – a world can comprise only one Smith. The reason for this is not hard to identify. Whether a pattern of mass-energy displacements is beautiful doesn't have anything to with its having certain origins – the Mona Lisa would be beautiful in any place or time. By contrast, whether a such a pattern Smiths or not is tied very closely to its conditions of origination. A qualitatively identical pattern that had different origins wouldn't be Smith – it would realize a Smith-duplicate, not Smith himself. But here the important point is that, quite plainly, whatever is the case with Smith – indeed, whether he exists or not – is fixed entirely by scientifically respectable lower-level properties. So it hard to see why we cannot countenance the property of realizing Smith or, therefore, of being identical with Smith.

The analysis given here applies to analytic as well as synthetic propositions, and atomic as well as molecular propositions.

§ I will argue that grammatical articulations *are* what distinguish word-heaps from sentences. I will also argue that grammatical articulations do not typically refer to anything. (In *some* cases they may, but not typically.) If grammatical articulations referred, then they would simply create big heaps where, previously, there were only small heaps. Suppose the inflections in (A) refers to entities x, y, and z. In that case, (A) wouldn't be a proposition; it

would be the heap: *Smith, Jones, the relation of punching, x, y, and z.* Referring to new things doesn't turn a heap into a proposition; so grammar articulations couldn't perform their unifying function by being referring terms. What grammatical articulations do is to indicate how to generate new properties on the basis of the properties introduced by the referring terms. (I am treating relations as properties – an n-place relation is a property had by n-tuples.) Consider (A). The referring terms – "Smith", "punches", "Jones" – pick out Smith, the relation of punching, and Smith. (For reasons I will give, Smith and Jones are themselves properties.) The purely *grammatical* properties of that sentence – the inflection on the verb, the relative positions of "Smith" and "Jones" – indicate how *new* properties are to be generated from those primitive properties. Those new properties – we will see soon enough what they are – suffice to encode the ordinal differences between *Smith punches Jones* and *Jones punches Smith.*

II. There are, ultimately, three conceptions as to what propositions are.

(i) Functions from worlds (or situations) to truth-values. So *that Plato snores* is the set of worlds where Plato snores or is a function that assigns truth to worlds where Plato snores, and falsity (or no truth-value) to other worlds.

(ii) States of affairs. So the proposition *that Plato snores* is identical with the event of Plato's snoring (or, perhaps, with a series of such events).

(iii) Arrangements of objects and properties. So *that Smith punches Jones* is some kind of arrangement of Smith, Jones, and the relation of punching, much as <2,6,8> is a structure consisting of three numbers

I believe that (i) and (ii) are false. I believe that (iii) is closer to the truth. But, first, I don't think it is quite correct. (As I will argue, *that Smith punches Jones* is not an arrangement involving Smith and Jones *per se*, but the properties of being identical with Smith and Jones, respectively.) But leaving aside this last point, (ii) is woefully incomplete. Suppose propositions *are* arrangements of individuals and properties. We still need to know two things. First, what is the relevant kind of arranging? We aren't talking about spatial arrangement. So what are we talking about? Second, how can an arrangement of objects and properties be *true or false*? The set (Smith, the relation of punching, Jones) isn't true or false? It doesn't matter how much ordinal structure we introduce: the ordered triple <Smith, the relation of punching, Jones> is no more true or false than its counterpart. We will deal with these problems in due course.

§ The Tractarian Answer

If I am reading him correctly, Wittgenstein gave a clear answer to the two questions we just asked. The sentence "Smith punches Jones" encodes an arrangement of objects and properties: Smith is related to Jones in some way, and the relation in question is of that of puncher to person punched. If reality comprises a state of affairs in which those things are arranged in that way, then the sentence is true. If not, it is false.

Suppose this view is right. In that case, the proposition *that Smith punches Jones* consists of Smith, Jones, and the relation of punching being arranged in a certain way; and they are arranged in the very way in which they would be arranged in the corresponding state of affairs. The sentence is isomorphic with the proposition. The proposition is isomorphic with state of affairs. That is why the sentence represents the state of affairs.

There is an obvious problem with this. If this view is right, then the thing meant by "Smith punches Jones" consists of Smith, Jones, and the relation of punching being arranged in the very way they are arranged in the corresponding state of affairs. How are they arranged in that state of affairs? Smith is punching Jones. So if the view in question is right, the proposition *that Smith punches Jones* consists of Smith standing in the relation of punching with respect to Jones.

But this means that the mere existence of the proposition guarantees its truth. If the proposition *that Smith punches Jones* consist of Smith's punching Jones, then the mere existence of that proposition guarantees that Smith is punching Jones. But the proposition *Smith punches Jones* exists regardless of whether Smith punches Jones. It is obvious that the way in which Smith, Jones, and punching are arranged in the proposition cannot be identical with the way they are arranged in the corresponding state of affairs.

At the same time, it does seem fair to say that *that Smith punches Jones* consists of Smith, Jones, and punching in some kind of arrangement. But, it would appear, the "picture-theory" seems not to give the right arrangement.

§ Are Propositions States of Affairs?

This naturally leads us to position (i): a proposition is a state of affairs. This is not a popular position, since there are two obvious problems with it. First, propositions are *true* or false. But a state of affairs is not true or false, and therefore isn't a proposition.[381] Second, (i) makes it impossible for there to be false propositions, unless one takes the heroic of measure of saying that there exist non-existent states of affairs.

Some (e.g. Barwise and Perry) have identified propositions with *possible* states of affairs or, at any rate, with sets of such things. This is not so different

from position (iii). Before we consider (iii), we should note one other reason why propositions cannot be states of affairs: this other reason is crucial to having a correct conception of what propositions are; it is also crucial to evaluating the Barwise-Perry position.

Consider (Ap). That proposition has a unique decomposition into minimal parts. It neatly decomposes into Smith, the relation of punching, and Jones. (It may have other ingredients, corresponding to tense-markers and the like. Let us ignore these for now.) Notice that, in that proposition, the relation of punching is completely distinct from Jones and from Smith. The proposition *connects* them, of course. But each of these things occurs in its purity: what occurs is not Jones' being tall or angry, but simply *Jones*. The proposition consists of three, neatly separated (though linked) entities.

In light of this, suppose that (Ap) is true. Consider the *state of affairs* that makes it true – imagine Smith punching Jones. Let S be that state of affairs. S doesn't have a structure even remotely like that of (Ap). S's minimal parts are definitely *not* the relation of punching, Smith, or Jones. First of all, Smith's act of punching is "adjectival" on Smith. It isn't as though Smith is here, and his act of punching is there. Smith's punching Jones is inseparable from Smith and Jones. Further, that act is by no means a *minimal* part of the states of affairs involved. It probably involved a number of subsidiary acts. (Even if it didn't, that doesn't affect our point. Consider the proposition *Beethoven composed the 9th symphony.* The "act" of composing involved innumerable deliberate subsidiary acts.) The same is true *mutatis mutandis* of Smith and Jones. They are not *minimal* parts of anything. They are constituted by innumerable events. If the state of affairs in question has minimal parts, those would be sub-atomic of some kind. But, I think, such events wouldn't really be parts of that event in quite the sense in which Jones and Smith are parts of the proposition *that Smith punches Jones.* It seems more natural to say that the state of affairs,

unlike the proposition, doesn't have *any* unique, correct decomposition. In any case, it doesn't have a decomposition even remotely like that of the corresponding proposition.

Also, Smith and Jones are not encountered *in their purity* in the relevant state of affairs. What occurs in the proposition appears to be simply *Jones.* Of course, the proposition goes on to ascribe a property to Jones. But Jones' occurrence in that proposition is unpropertied. What occurs is not Jones' being tall or Jones' being sweaty - just *Jones.* But in any state of affairs, even one that involves Jones' floating in a void, what occurs is not just *Jones*: it is a state of affairs involving Jones. What occurs in reality is Jones' being a in a certain place, having a certain height and weight, animated by certain attitudes or beliefs. There is no Jones *simpliciter.* There are states of affairs involving Jones. Jones is inseparable from them. It is not as though there is *first* a Jones and *then* a state of affairs is constructed out of Jones. For Jones to exist *is* for a state of affairs of some kind to exist. Jones is not accurately thought of as an isolable constituent of states of affairs. Jones *simpliciter* – the Jones that doesn't occur in a state of affairs – is a pure abstraction. What occurs in reality is not Jones *simpliciter*; what occurs are Jonesy states of affairs. There can no more be "free" occurrences of Jones – occurrences of Jones in isolation of a state of affairs -- then there could be free occurrences red. For these reasons, I think that *metaphysically* Jones is a property of states of affairs. As for what exact kind of property he is, that depends on your other metaphysical views. If you believe the essentiality of origins, then a state of affairs is Jonesy if it has a certain history. If you have a more theologically inclined view, you might say that a state of affairs is Jonesy if it is animated by the right kind of immortal soul. All that is important here is that Jones must be seen as a property of states of affairs, lest we run into insuperable metaphysical problems.

§ It is clear that the *articulations* one finds in a proposition are radically different from those found in the corresponding states of affairs. Ian Hacking refers to this as the "articulation problem". In the proposition *Socrates is tall*, we have one constituent corresponding to Socrates and a totally different one corresponding to tallness. But Socrates cannot be separated from his tallness. Even if you cut off Socrates' legs, you are not separating him from his height: you are only giving him a new height, and he will be as inseparable from his new height as he was from the old one. Propositions separate out what, in reality, are balled together. Propositions and states of affairs have entirely different decompositions. Various thoughts and atoms are constituents of Socrates. But no thoughts or atoms are constituents of the proposition *that Socrates is tall.* So propositions are not states of affairs, and they are not things that are even remotely *like* states of affairs.

However we analyze propositions, we must make sure that the things with which we identify propositions have the same decomposition as the propositions themselves. It is practically a datum that Socrates is not only a constituent of the proposition *that Socrates is tall*, but a *minimal* and *discrete* part of it. The idea that propositions are states of affairs is inconsistent with that idea. For given any state of affairs in which Socrates is involved, he is absolutely not a minimal part of it, and he is absolutely not neatly separable from it. This fact, I believe, vitiates the idea that a proposition is a set of states of affairs or, what is not so different, a set of worlds. A set of states of affairs has whole states of affairs as ultimate constituents: Socrates himself is no ultimately constituent of such a set. And any given state of affairs cannot possibly have Socrates as an ultimate constituent.

§ This brings us to approach (iii) Consider the set of worlds where Socrates is tall. Socrates isn't a constituent of that set in remotely the sense in which, by

all appearances, he is a constituent of *that Socrates is tall*. Now consider a function that assigns truth to each of the worlds just mentioned. The point we just made still applies. On face of it, Socrates isn't a constituent of that function in the sense in which he is a constituent of *that Socrates is tall.*

Of course, attempts have been made to reintroduce constituent-structure into that function – to make it more like the corresponding propositions. Whether these attempts succeed is a question for another time. Even if they do, the possible-worlds approach is still a failure. If a possible world is a set of propositions, then (ii) is viciously circular. If a possible world is a concrete alternative world, then (ii) is committed to all manner of dubious metaphysical and epistemological views. (There is no point in going through them here: they are well documented.) Given this, it would probably be better to abandon (ii).

§ A word of clarification is in order here. I think that it is *true* to say that propositions are functions from worlds to truth-values. But such a statement is unacceptable *if taken as an analysis*. There are true, and even informative, statements that cannot serve as analyses. Permit me to explain my meaning. First of all, there is surely nothing wrong with talking about hypothetical worlds – such talk is innocent. Let w be a hypothetical world. Some propositions will be true in w and some won't. Let P be a proposition that is true in w. Since P is true in w, we can naturally think of P as associating w with the truth-value *true* (or the property of truth). Let w* be some other hypothetical world, but suppose that P is false in w. P associates falsity with w*. And so on, for any other hypothetical world one might consider. So it is certainly true that propositions associate truth-values with hypothetical worlds, and can thus be regarded as functions from worlds to truth-values.

But propositions *are* functions from worlds to truth-values in the sense of "are" that indicates *predication.* All and only propositions have the property of

assigning truth-values to worlds. But if we *analyze* propositions as functions from worlds to truth-values , then our statement is either viciously circular (if the worlds in question are treated as propositions) or is wrong for other reasons (if the worlds are treated as concrete objects).

A comparison might be in order. It may well be *true* that x is the same event as y iff x and y stand in the same causal relations, i.e. if everything that causes x also causes y and if anything that x causes is also caused by y. And such a statement is not only (quite possibly) true, but also informative. But taken as an *analysis* of the concept of an event, that statement is circular: after all, causes and effects ere themselves events.[382]

Here is another example. Nothing that is not a proposition entails anything; and anything that is a proposition entails something. So

(*) all and only propositions stand in entailment relations.[383]

is true. Further (*) is an *informative* statements. But it is not acceptable as an *analysis* of the concept of a proposition. Anything entailed by a proposition is itself a proposition. So (*) amounts to:

(**) all and only propositions entail other propositions.

Taken as an analysis, (**) is viciously circular. I believe the statement *propositions are functions from worlds to truth-values* is informative and accurate. But it is viciously circular if it is taken as an analysis. I believe that PWS is a giant projection of a failure to distinguish the conditions that must be met to provide an analysis of a concept from those that must be met to make a true statement about a concept.

(i) is an out and out failure. (ii) is either viciously circular or is extremely dubious. Further, it is at best an open question whether a set of worlds or a function from worlds to truth-values could have a decomposition at all like that of a proposition. Given this, we should, I believe, try to make some variant of (iii) work.

III. Now it is time to give a positive solution. Let us start with some basic points about propositions. To a first approximation, a proposition is a property of a world. Suppose that *Plato snores* is true in w and false in w*. In that case, w has a property that w* does not. w comprises certain states of affairs that w* does not. The space-time manifold in w has a ripple in it *not* had by the space-time manifold in w*. Kenneth Taylor (1998) speaks of states of affairs as "ripples in the quantum". For Socrates to be snoring is for the quantum to be rippled or indented in a certain way (with the possible qualification that those ripples have to have certain origins).

Being tall or red is a way that individuals can be. Thus tallness and redness are properties of individuals. A proposition is a way a world can be. It is thus a property of a world. This is why David Lewis (1986) explicitly identifies propositions with properties of worlds. For the proposition *Socrates snores* to be true in w is for w to be a certain way – it is for w's quantum to have certain ripples in it. That proposition is thus a property, and that proposition is true in a world exactly if that proposition is *instantiated* in that world. We can thus identify truth with instantiatedness. A proposition P is a property of a world, and P is true exactly if that property is instantiated.

Here we should deal with what would appear to be a major problem with what we've said so far (to avoid confusion, I'm going to put the objection in the mouth of an imaginary interlocutor):

Let p be the property of being a world where

(*) *Plato punched Socrates*

is true. Let w be a world where the proposition *Plato punched Socrates* is true, and let w* be a world where that same proposition is false. w has p. w* does not. So it is obvious that, for some property x, *Plato punched Socrates* is true in a world exactly if that world has x. This point generalizes without limit: given any proposition P, there is some property P* such that P is true in a world exactly if that world has P*.

But here is the problem. The proposition seems to be more basic than the corresponding property: the property must be understood in terms of the property; and the property's being instantiated must be analyzed in terms of the proposition's being true. This spells the doom of your analysis: for you are trying to analyze the proposition in terms of the property, and the proposition's being true in terms of the property's being instantiated.

For (*) to be true in w, what property must (*) have? It must have p. But what is p? It is the property of being a world such that, in that world, (*) is true. So (*) is more basic than p. p is defined in terms of (*). This means that we cannot analyze (*) in terms of p; and we cannot analyze what it is for (*) to be true in terms of p's being instantiated. On the contrary, we must analyze p in terms of (*), and we must analyze p's being instantiated in terms of (*)'s being true.

This point generalizes without limit: what we just said about (*) is true of any proposition. Consider the proposition

(**) *penguins don't fly.*

Obviously there is some property P such that (**) is true in a world w exactly if w has P. P is the property of being such that (**) is true in it. But we cannot analyze (**) in terms of P, and we cannot analyze (**)'s being true in terms of P's being instantiated. On the contrary, we must analyze P in terms of (**), and P's being instantiated in terms of (**)'s being true. So your account seems doomed to vicious circularity.

I think that this point is entirely wrongheaded. (To simplify discussion, let us leave momentarily aside analytic propositions.) A proposition is never *just* true. It is always true in virtue of something about the world: mass-energy is distributed in the right way; sentient beings are having the right thoughts or feelings. A true proposition registers some fact, and the truth of a proposition supervenes on something more basic.

Consider (*). It is not as though Plato *just* punches Socrates. Plato's punching Socrates supervenes on facts of a much more fundamental kind: Plato punches Socrates in virtue of the fact that atom a moves in this way at this time, atom b moves that way at that time, and so on. Given two worlds where the same elementary particles move in the same way, either (*) is true in both of them or in neither. So the truth of (*) supervenes on the way in which mass-energy is distributed in the world. That distribution could obviously be described without mentioning Plato or Socrates or the relation of punching. Given any one of the electron-jumps involved in Plato's punching Socrates, that electron-jump could obviously be described without mentioning Plato or Socrates. The same applies to each of the micro-events involved Plato's punching Socrates. Obviously the totality of those micro-events could (in principle) be described without mentioning Plato, Socrates, or the relation of punching. That totality is plainly *more* basic than (*). If (*) is true in a world

w, that fact supervenes on the way in which space-time is wrinkled in w. Those wrinkles *underlie* the truth of (*); and they *can* plainly be described without mentioning (*). Thus the property of having those wrinkles is more basic than (*), and the property of having those wrinkles is more basic than the concept of (*)'s being true. So there is no circularity in analyzing (*) in terms of the property a world has if it is wrinkled a certain way, and there is no circularity in analyzing (*)'s being true in terms of a world's having the right wrinkles.

There is more to say in this connection. As we've noted, given any two worlds where mass-energy is distributed in the same way – any two world's whose quanta are rippled in the same way – (*) is true in both of them or in neither: so the truth of (*) supervenes on other facts. But the converse is not true. There can be two worlds where (*) is true but that otherwise differ enormously. There are many different ways that Plato can punch Socrates, and there are many different ways that the surrounding circumstances can vary. So whether (*) is true is fixed by how the quantum is rippled in w. But how the quantum is rippled in w is by no means fixed by whether (*) is true in that world. This puts it beyond doubt that, if (*) is true in a world w, that is in virtue of w's having some property that is *more basic* than (*): it is in virtue of facts about the basic constituents of w – facts about the way the quantum is rippled in w. The property that a world has if (*) is true in it is thus less basic than the property that a world has if the quantum is rippled the right way in it. So, contrary to what the objector says, there is some property P such that P is more basic than (*), P's being instantiated is more basic than (*)'s being true, and such that (*)'s being true is fixed by P's being instantiated (but not *vice versa*). Thus our account is *not* guilty of vicious circularity, despite what the objector says.

These points generalize. Consider any empirical proposition p. p is not *just* going to be true in w. It is going to be true in virtue of some fact about the

basic constituents of w. p will be true in w in virtue of the fact that, as we put it, the quantum in w is rippled a certain way. The property of having that rippling will be *more* basic than p, and a world's having the right rippling will be *more* basic than p's being true. So I think the objector's point is misguided.

Let us sum up. Consider the proposition *Jones punches Smith.* This is true in w iff the "quantum" in w is "rippled" a certain way – if w comprises displacements of mass-energy of a certain kind. So the proposition *that Jones punches Smith* can be thought of as a property of worlds, and that proposition/property is true in a world w iff it is instantiated in w.

IV. But the story just told isn't quite accurate. There are a couple of reasons for this. Let p be the property that a world must have if, in it, Jones punches Smith. p will be instantiated in w exactly if, in w, there is a certain pattern of mass-energy displacements. But exactly these mass-energy displacements will suffice for the truth of infinitely many other propositions. Obviously they will suffice for the truth of every analytic proposition. (Since such propositions are true under any circumstances, it follows vacuously that whatever happens in a world is sufficient for their truth.) But those same displacements will make true *somebody punches Smith* and *Jones punches somebody* and *somebody punches somebody* and *somebody does something intentionally.*

We could try to get around this. If a mass-energy displacement is enough to make true *Smith punches Jones*, it is enough to make true *somebody punches Jones*, but the converse doesn't hold. So perhaps we could use this asymmetry to distinguish the kinds of properties associated with the patterns of mass-energy displacements.

But this maneuver won't always work. Consider the proposition *x is a circle* and *x is a closed two-dimensional figure of uniform curvature.* Surely these are different propositions; and surely in any possible world, any mass-energy

displacement sufficient for the truth of the one is sufficient for that of the other. Of course, some will deny that these are different propositions. But that denial has little independent motivation – it seems an ad hoc way of ameliorating a commitment to a degenerating research program.

Also, the Lewisian picture seems to make analytic propositions disappear. If a proposition is a property of a world, then what a proposition tells you is nothing more and nothing less than how a world with that property is different from a world without it. An analytic proposition is one that all worlds have. So, on Lewis' model, analytic propositions don't tell you anything: they might as well not exist.

The problem with the view we've been considering is one we've already encountered. Propositions have decompositional structure. A proposition is not specified merely by its truth-conditions. It is necessary also to give its decomposition The Lewisian analysis just considered doesn't accommodate the internal or decompositional differences between

(i) *Alpha is a figure whose perimeter coincides with the class of all points equidistant from a given point in a plane*

and

(ii) *Alpha is a closed two-dimensional figure of uniform curvature.*

So it is not enough to say that a proposition is a property of a world. This obviously relates to the fact that analytically equivalent propositions can convey very different information and, what is a consequence, that analytic propositions *do* typically convey information. At the same time, the view just considered contains some profound truths, and our view will be more a

development than a rejection of it.

Another illustration may help. Again consider:

(Ap) *Smith punches Jones.*

In this world, if Jones and Smith are to exist, innumerable sub-atomic, atomic, chemical, molecular, cellular…events must occur. In *any* world where Jones and Smith exist, their existence supervenes on the existence of lower-level phenomena. The same is true of the relation of punching: if there is a punching, that supervenes on many other events. So no state of affairs that makes (Ap) true has have Smith or Jones or the relation of punching among its ultimate constituents. Given this, suppose we identify (Ap) with some property P such that (Ap) is true in w exactly if P is instantiated in w. In any world, an instance of P will make true many propositions whose ultimate constituents are *not* Smith or Jones – whose ultimate constituents are tiny displacements of mass-energy. This makes it hard to believe that P itself has a unique decomposition whose ultimate constituents are specifically Smith, Jones, and the relation of punching. If Jones manages to punch Smith, that is only by the grace of innumerable subvenient states of affairs. So P's being instantiated demands the occurrence of subvenient states of affairs having certain properties. So it is individuative of P, and thus essential to it, that whenever it is instantiated, things *more basic than* Smith, Jones, and the relation of punching are instantiated. For this reason, it cannot be said with confidence that P has a unique, ultimate decomposition into Smith, Jones, and the relation of punching. On the contrary, it seems as though those three things are *not* among its ultimate constituents. This makes P unsuited for identification with (Ap).

All of this assumes that properties themselves even *have* constituents. It is

clear that instances of properties have constituents. An instance of the property *chair* has legs and a back. But it isn't really clear if that property itself has constituents. Even if properties do have constituents, it isn't clear if they have them in the sense in which propositions have constituents. Perhaps the property *water* is composed of the property *hydrogen* and the property *oxygen*. But are the last two properties constituents of the first in the very sense in which *Smith* is a constituent of (Ap)? This is, at best, an open question. I will argue that the answer is "no".

Let us sum up. It is pretty clear that Smith and Jones and the relation of punching are constituents of (Ap). It is also clear that (Ap) has a unique decomposition into minimal parts, and that the three things just mentioned are among those minimal parts. But it is *not* clear if P has an even remotely comparable decomposition. It isn't even clear if P has *a* decomposition in the sense in which (Ap) does. Also, P's being instantiated, while sufficient for the truth of (i), is also sufficient for the truth of infinitely many distinct propositions. So it does not seem right to *identify* (i) with P.

§ There is one other desideratum that any theory of propositions must satisfy. Consider the following propositions:

(Ap) *Smith punches Jones.*
(2) *Smith punches somebody.*
(3) *Smith punches everybody.*
(4) somebody punches somebody.
(5) Fred punches nobody.
(6) Fred punches Jones.
(7) Nobody punches Jones.

It is pretty clear that all of (Ap)-(&) have a constituent in common, namely: *__punches__* or *x punches y.*

It is pretty clear that (Ap)-(3) that have a constituent *in common*, namely: *Smith punches__* or *Smith punches x.*

It is pretty clear that (Ap)-(3) that have a constituent *in common*, namely: *Smith punches__* or *Smith punches x.*

It is pretty clear that (6)-(7) that have a constituent *in common*, namely: *__punches Jones* or *x punches Jones.*

The naïve thing to say is that (Ap) has *three* constituents: one corresponding to Jones, one corresponding to Smith, and one corresponding to the relation of punching.

But that would overlook the fact that (Ap) has the component *x punches Jones* and also the component *Smith punches x.* One of (Ap)'s components seems to correspond to the property of being a thing x such that Smith punches x. Another seems to correspond to the property of being a thing x such that x punches Jones. So (Ap) has several constituents. It has one corresponding to Jones; one corresponding to Smith; one corresponding to the relation of punching; one corresponding to the property of being a thing x such that Smith punches x; one corresponding to the property of being a thing x such that x punches Jones; and, finally, one corresponding to the result of putting all of these things together.

Frege argued that the proposition *Smith punches Jones* is built out the relation *x punches y*, and the functions *x punches Jones* and *Smith punches x.* The point we just made obviously has a certain similarity to Frege's. (In fact, I will argue, they are the same point.)

§ Here is how I would propose to accommodate all of these facts. (Ap) is a *set* S of properties, *one* of whose members is P. But P is not the only member of S. What are its other members? My answer to this question will initially be alarming.

Let us start with terminology:

Ps: the property of being identical with Smith.
Pj: the property of being identical with Jones.
Psx: the property of being a thing x such that Smith punches x.
Pxj: property of being a thing x such that x punches Jones.
Pp: the relation of punching.
P: as previously defined.

I propose that (Ap) is a set S whose members are exactly the properties just listed. And I propose that for (Ap) to be *true* is simply for all of its members to be instantiated.

For reasons already given, our including P on the list of relevant properties does *not* constitute a case of vicious circularity.

Let me now give the motivation for this analysis. First of all, what circumstances is (Ap) true? The following properties must be instantiated: the relation of punching (i.e. the relation *x punched y*); the property of being a thing x such that Smith punched x; the property of being a thing x such that x punches Jones; the property of being identical with Smith; the property of being identical with Jones; and finally P.

If any one of these properties is not instantiated, then (Ap) is not true. For example, if the property *identical with Smith* is not instantiated, then there is

no Smith; and there is therefore no Smith to punch Jones. So for (Ap) it is *necessary* that each of those six properties be instantiated.

It is also *sufficient.* Suppose that each of those properties is instantiated. In that case, there is a Smith, a Jones, a case of somebody's punching somebody, a case of Jones' being punched, a case of Smith's punching somebody, and a case of Smith's punching Jones. (The instantiating of P guarantees the occurrence of the last of these.)

There is more to say. The proposition (Ap) has a discrete parts corresponding to Smith, Jones, the relation of punching, the property of being a thing x such that x punches Jones, and the property of being a thing x such that Smith punches x.; and (Ap) also has a (maximal) discrete part corresponding to the result of putting these together.

The same thing is true of S. S has one member corresponding to Smith, one corresponding to Jones, one to the relation of punching, and so on. So S has a breakdown parallel to that of (Ap).

So here is what we have thus far. The instantiating of those six properties is both necessary and sufficient for the truth of (Ap). Further, for each discrete constituent of (Ap), there is one property on that list; and for each property on that list, there is one constituent of (Ap).

Here we should address a worry one might have about our analysis? The instantiating of P *by itself* is necessary and sufficient for the truth of (Ap). So why not just say that (Ap) is identical with P and that (Ap)'s being true is identical with P's being instantiated?

The reason is that P does not decompose in the same way as (Ap). If we are claiming that some thing x is identical with (Ap), we must at all costs make sure that x has Smith, Jones, and the relation of punching, and so on, as discrete constituents. If we identify (Ap) with P alone, it is by no means clear if that burden is met. But if we identify S with the set containing exactly the

six properties listed above, then, first, (Ap) is true exactly if every member of S is instantiated and, second, the set just described has the same decomposition as (Ap): like (Ap) that set consists of Smith, Jones, the relation of punching, and so on.

As we said earlier, any analysis of propositions must accommodate the fact that different propositions can have common constituents. Our proposition has that virtue. Consider the proposition:

(PJ) *Plato punches Jones.*

Here is what we pre-theoretically believe about the relation between (Ap) and (PJ). They have certain constituents in common: one of these constituents corresponds to the relation of punching; another to the property *x punched Jones*; and another to Jones. For various reasons, I think we should hold onto this pretheoretic intuition. The facts relating to the "systematicity" and "compositionality" of thought accord very well with this intuition; they would not accord so well with its negation.

Now remember what we said about (Ap). Suppose that we apply exactly the same analysis (mutatis mutandis) to (PJ). What we end up with this is this. (PJ) is a set S* whose members are exactly the following properties:

Pp: the property of being identical with Plato
Pj: the property of being identical with Jones.
Ppx: the property of being a thing x such that Plato punches x.
Pxj: property of being a thing x such that x punches Jones.
Pp: the relation of punching.
P*: just like P (mutatis mutandis).

And (PJ) is true exactly if every member of S* is instantiated.

S and S* have certain members in common: Pj; Pxj; and Pp. These members corresponding to the constituents that we pre-theoretically believe (Ap) and (PJ) to have in common, namely: Jones; the property of punching Jones; and the relation of punching. So our analysis *accommodates* the fact that (Ap) and (PJ) have exactly those constituents in common. For exactly similar reasons, it *accommodate* the fact that both those propositions have one constituent in common with *Fred punches Barney*; and that one, but not the other, has a constituent in common with *Fred punches Jones.* For the same reason (mutatis mutandis), given *any* two propositions that pre-theoretically believe to share some constituent c, our analysis is consistent with the fact that they share that constituent.

§ I wish to stress one point. We identified (Ap) with a set S containing all and only the six properties listed above. But it might seem that our inclusion of P on that list dooms our analysis to vicious circularity. P is the property of being a world where Smith punches Jones. So P is defined in terms of the proposition *Smith punches Jones.* But we are analyzing that very proposition in terms of P. So our analysis is circular.

But there is no circularity. We need some way to *identify* that property. And the easiest way to do so is in terms of a proposition – in terms of (Ap). When we *describe* an event, we typically use a sentence and, therefore, a proposition. (I say “typically” and not “always” because an event can also be described through some analogue medium, like a motion picture.) We say “Smith punched Jones” or whatnot. But consider the mass-energy displacements which realize that event; consider the over-all mass-energy storm constituted by those various displacements. That storm doesn’t have remotely the kind of structure had by any of the propositions which describe it.

It isn't as though Jones can *just* exist. Jones' existence supervenes on that of innumerable other states of affairs. In fact, the entity *Jones* is really a kind of abstraction from these various states of affairs. In the proposition, *Jones* is separate from his various attributes. Jones isn't separate from his tallness, even though the proposition *Jones is tall* has one constituent corresponding to Jones and a discrete one corresponding to his tallness. The same is true (mutatis mutandis) of Jones' being punched Smith. the proposition *Smith punches Jones* has (at least) three discrete constituents. But the event thereby described – the storm of mass-energy displacements -- is a seamless whole; in any case, so far as it does have articulations, they are not remotely like those of that proposition.

There are, in principle, different ways to identify that storm. It can be indicated by ostension. In fact, that is how it *was* identified by those who witnessed it. ("Harry, did you see *that*? Wasn't *that* astonishing? I never though Smith was such a cad.?") Of course, that event can also be indicated by description. When we *describe* an event, we typically use a sentence and, therefore, a proposition. (We don't have to use propositions: we could also use analogue media like motion-pictures or painting.) Under the circumstances, it was necessary for us to *identify* that event by way of a sentence and, therefore, a proposition. But we didn't *have* to so identify it. In principle, it could have been identified through an ostension or a drawing or a motion-picture.

Now consider the *kind of* mass-energy storm we've just been discussing. In any world where (Ap) is true, there is an instance of that kind. To *identify* that kind, we use a sentence and, therefore, a proposition: we say, for example, "it is the way of being wrinkled that is common to all and only those worlds where *Smith punched Jones* is true." But we are using that proposition "referentially", not "attributively"; we are using it a means of identification only. We are analyzing (Ap) in terms of the corresponding *kind* – not in terms

of the proposition used to *pick out* that kind. We use the propositioin *only* as a means of identification. It is the thing identified, not the means of identifying it, that enters into our analysis. So our analysis is not circular.

But there is a possible objection to this:

You say: "consider *the* kind instantiated by all and only the mass-energy storms realizing the event of Smith's punching Jones." But that storm instantiates all manner of different kinds. Qualitatively different sorts of storms can realize a case of Smith's punching Jones. Smith and Jones can be located in any number of places and any number of times. The punch can be thrown in all kinds of different ways. And so on. So when you talk about *the* kind of mass-energy storm A given storm instantiates all manner of different kinds. So which kind are you talking about? You are talking about the kind instantiated by all and only those storms sufficient for the truth of *Smith punched Jones.* So that proposition is not *merely* a way of picking out a property; it enters into the very definition of the right kind of property; so your analysis *is* circular.

(Ap) is a contingent proposition. In this context, let us focus on such propositions. A proposition doesn't *just* hold. It holds in virtue of some fact – some state of affairs, some mass-energy displacement (or, more likely, consisting of such displacements). If a proposition is true in a world, there is something that makes it true – something that anchors it. Propositions aren't free-wheeling. A true proposition is anchored to the world by a non-propositional entity.

There are a couple of corollaries. First, if a proposition is true in a world, that imposes constraints on the facts constituting that world. Given that *Smith*

punched Jones is true in w, there are some limits on how the underlying facts can be. There are some limits on what the underlying mass-energy displacements can do.

A consequence is that if a given proposition is true in two worlds, then those worlds have something in common *in virtue of which* that proposition is true in both of them. It is not a *brute* fact that they are both true; the commonality is anchored in some extra-propositional commonality. If we say that there is no such commonality, we are saying that the underlying facts can be vary without limit; we are saying that the truth of that proposition imposes *no* limits on the underlying facts. The mass-energy displacmenets can occur in any way at all. But if a proposition is true, that plainly *does* impose limits on these underlying facts.

So if *Smith punched Jones* is true in two different worlds, that is in virtue of some *extra*-propositional commonality. It cannot be a brute fact. The proposition *Smith punched Jones* is obviously a useful way of *identifying* that commonality. But that commonality *underlies* the fact that *Smith punched Jones* is true in both worlds. That fact cannot *constitute* the commonality. Otherwise, we strip the proposition of any anchoring in extra-propositional reality; and we deprive true propositions of the power to impose limits on the how the underlying facts may be. So if *Smith punched Jones* is true in two different worlds, that indicates some more *fundamental* commonality. There is some more property P such that both worlds have P and such that it is a *consequence* of their having P that *Smith punched Jones* is true in each. The description “property had in common by all and only those worlds where *Smith punched Jones* is true” is a useful way of identifying that property. But the property is more fundamental than the corresponding proposition. And what enters into our analysis is the property itself, not the description of it. That is why our analysis is not circular.

§ One worry about our proposal is this. S doesn't have *Smith* has a member: instead, it has the property of being identical with Smith. By contrast, (Ap) has Smith, not the property of identical with Smith, as a constituent. Therefore, S cannot be identical with (Ap).

There are (at least) two different ways we can respond to this. I don't think that either is more correct than the other.

For a long time, semanticists have insisted that Smith himself is a constituent of the proposition *Smith punches Jones.* But when we say such a thing, we are making an enormously theoretical claim. We don't have any direct understanding of what propositions are. When we say that make some claim about what the "constituents" of propositions are, it is to accommodate various data (about modality, analyticity, cognitive content, and the like). For example, we know that *that Smith punches Jones* couldn't possibly be true if there were no such thing as Smith. So this leads us to say that Smith himself must be a "constituent" of that proposition.

But if there should turn out to be some other way to accommodate this modal datum, then there would be no need to insist that Smith himself is a constituent of that proposition. We have identified (Ap) with S. S's being instantiated occurs exactly if (Ap) is true. So S's modal properties are in lock-step with (Ap)'s. Thus if we were to analyze (Ap) as being identical with S, that would be consistent with the relevant modal data. Since that analysis accommodates the relevant intuitions about modality, we (to that extent) have grounds for thinking it is the property of being identical with Smith, and not Smith *per se*, that is a constituent of (Ap). Our analysis does *not* accommodate the view that Smith himself is a constituent of (Ap). But *that* view is an entirely theoretical one; it is *not* a datum. It is surely not a *datum* that Smith is

a constituent of that proposition. So our analysis cannot be thrown out on the grounds that it doesn't accommodate *that* view.

An analysis is meant to accommodate all the relevant *data* – not al the relevant theoretical superstructures that have grown up over that data. We can't say that Einstein's theory is wrong simply because it conflicts with the "datum" that Newton's theory is right. The latter is not a datum. We cannot say that our analysis of propositions is wrong simply because it makes the property of being identical with Smith, and not Smith himself, be a constituent of certain propositions. The idea that Smith himself is such a property is a bit of theoretical superstructure; it doesn't have the same standing as any datum. And our theory is consistent with the relevant data.

§ There is another way to respond to the problem in question. We *could* say that our analysis *is* consistent with the idea that Smith is a "constituent" of (Ap). The idea would be this.

When we say that Smith is "constituent" of (Ap), that is really a metaphorical way of saying that the truth of (Ap) depends on some fact about Smith: some kind of relation of logical *dependence* holds between (Ap) and Smith. The sense in which Smith is a "constituent" of a proposition is entirely distinct from the sense in which something is a member of a set. The *membership* of S is not identical with the constituency of (Ap); and in order to do justice to the constituency of (Ap), we *must* assign members to S that are not constituents of (Ap). It is precisely *in order* to accommodate the fact that Smith is a constituent of (Ap) that we must make the property of being identical with Smith a constituent of S.

Let me clarify these obscure remarks. Suppose we take it as a given that Smith himself must be a constituent of (Ap). (I myself don't think that is a "given" at all. But let us leave that aside.) *By itself* that doesn't mean that our

analysis is wrong. True – our analysis identifies (Ap) with some set S. True – S has as a member not Smith himself, but rather the property of being identical with Smith.

But it still doesn't follow that our analysis is inconsistent with the idea that Smith is a "constituent" of S. After all, we haven't yet analyzed the "constituency" relation yet. When we say that object O is a "constituent" of proposition P*, we are making a theoretical claim. We certainly cannot *see* that Smith is a constituent of some proposition the way we can *see* that a certain brick is a constituent of a certain house. When we talk about propositions, the term "constituent" is obviously being used in a metaphorical sense. Smith isn't a constituent of a proposition in the sense in which a leg is a constituent of table. Propositions are abstract objects. They don't have "constituents" in any straightforward sense. So even if we take it as a datum that Smith is a constituent of (Ap), we cannot say that our analysis is wrong *until* it is has been said what it is for something to be a constituent of a proposition. Once that has been said, *then* it can be determined whether our analysis is wrong.

But I believe that the concept of "constituency" can be delineated in a way that reconciles our analysis with the aforementioned "datum." Suppose that (Ap) is true in some world w. According to our analysis, that means that the members of S are all *instantiated* in w. (Ap) is true exactly if all the properties in S are instantiated. One of those is the property of being identical with Smith. So, if our analysis is right, then (Ap)'s being true consists in various properties being instantiated, and Smith will be identical with one of the resulting *instances.*

So Smith is a "constituent" of (Ap) in the following sense. (Ap) is a set S of properties. For (Ap) to be true in w is for those properties to be instantiated in w. The property of being identical with Smith is one of those properties. So Smith is identical with one of those instances. Smith is a "constituent" of (Ap)

in the sense that he is an *instance* of one of the properties that is a member of S.

In general, the constituency relation would amount to this. Let P* be an arbitrary proposition. P* is identical with some set S* of properties. O is a "constituent" of P* exactly if O is an instance of one of the properties that is a member of S*.

§ Let us take stock. Let (Ap) be the proposition *that Smith punches Jones.* According to our analysis, (Ap) is identical with a set of properties S. S has for its members various properties. One of the members is the property of being identical with Smith. Smith himself is not one of those members.

Given this, there are two ways we can deal with the widely held view that Smith is a "constituent" of (Ap). We can say:

(*) That view is wrong. It turns out that what is a member of S is not Smith himself, but the property of being identical with Smith.

The idea that Smith is a "constituent" of (Ap) is not a datum; it is a hypothesis – a piece of theoretical superstructure. One cannot deny that this table-leg is a constituent of this table: it is a datum; we *see* that it is true. But one cannot just *see* that Smith is a "constituent" of (Ap). The position that he is such a constituent is not a datum, but is rather a theoretical way of *dealing* with various data. Among these data are modal data, e.g. the datum that (Ap) couldn't possibly be true in a circumstance where there were no Smith. Our analysis of propositions accommodates that datum. And our analysis accommodates it by *denying* the theoretical (and obscure) view that Smith is a constituent of (Ap). So, methodologically speaking, we have good reasons for *denying* the presumption that Smith is himself a constituent of (Ap).

The other position we can adopt is this:

(**) Smith *is* indeed a "constituent" of (Ap). Our analysis is consistent with that. Indeed, our analysis makes it clear what it is for Smith to be a constituent of (Ap). (Ap) is identical with a set of properties S. For (Ap) to be true is for all the members of S to be instantiated. One of those properties is that of being identical with Smith. In any world where (Ap) is true, Smith is an instance of one of the properties that is a member of S. Smith's being a "constituent" of P consists in its being the case that, in any world where (Ap) is true, Smith's is an instance of one of the properties that is a member of (Ap). In general, for O to be a constituent of a proposition P* is for it to be the case that, in any world where P* is true, O is an instance of one of the properties that is a member of P*.

I think that (*) and (**) are really the same view, the only differences being verbal.

§ Let us sum up. (Ap) is true exactly if all the members of S are instantiated. Further, (Ap) and S have the same decomposition. Therefore we might as well identify (Ap) with S and (Ap)'s being true with S's being such that its members are all instantiated.

At this point, there might seem to be a problem for our analysis. We just said that (Ap) and S have *exactly* the same decomposition: and surely they must, if S is to *be* (Ap). But S appears to have a constituent that Ap does not have. S comprises P; but (Ap) comprises only Smith, Jones, and the relation of punching – there is no P. So S and (Ap) have different constituents and are not

identical.

I think this is false. I think that (Ap) *does* have a constituent corresponding to P. Smith, Jones, and the relation of punching are not the only constituents of (Ap). That proposition also has a *complex* constituent. Consider the verbal representation of that proposition: “that Smith punches Jones”. The words “Smith punches Jones” seem to correspond to a single entity, albeit one that is, in some sense, composed of Smith, punching, and Jones. It is this complex entity gives (Ap) its truth-conditions. Smith, Jones, and the relation of punching contribute to (Ap)’s truth-conditions only by way of their involvement in this complex. So while it is true that one of S’s constituents is something other than Smith, Jones, and punching, the same is true of (Ap).

§ One more application of our analysis might of use. Consider the proposition

(#) *Plato snores.*

Let w be a world where (#) is true, and let w* be one where (#) is false. w is different from w*. The quantum in w is rippled differently: in w, there are events having a certain character and having certain origins; in w*, there are no such events. Let P be the property a world has iff its quantum is rippled in that way. (#) is true in a world exactly if P is instantiated in that world. This might induce one to say that (#) is *identical with* P and that for (#) to be true is identical with P’s being instantiated.

But as it stands, this position is not quite acceptable. (#) has a unique, ultimate decomposition; and Plato and the property of snoring are uncontroversially among those ultimate constituents. (Whether they are *all* of them is another matter: I think not, for reasons I’ll soon give.) Given this,

suppose we identified (#) with P. In that case, if our identification is to be correct, then P must have the same decomposition as (#). But it is by no means clear whether this condition is met. If P is instantiated, let us say that there is a P-state of affairs. A P-state of affairs necessarily supervenes on other, lower-level states of affairs: states of affairs involving tissues in somebody's nasal passages, various interactions among air-molecules, respiratory and neurological processes, not to mention innumerable sub-atomic processes. Of course, the exact realization of any P-state of affairs may vary from world to world (or even within a world). But what is *not* thus invariant is that if there is a P-state of affairs, that is wholly in virtue of the existence of innumerable lower-level states of affairs meeting certain broad requirements. (It may not matter *exactly* how the air-molecules are distributed. But it matters that they go through *some* kind of passage in a body meeting *certain* requirements, with the result that noise of a *certain* kind is produced: within these limits, there is enormous latitude . But there are limits.) So a necessary condition for P's being instantiated is that various lower-level phenomena are instantiated. The necessary conditions for a property's being instantiated presumably reflect facts about the structure of the property. Given this, it becomes at best an open question whether P decomposes neatly into Plato and the property of snoring. There is the distinct possibility that, if it decomposes at all, it decomposes into things much more basic than those two. Also, as we said before, there is no clear sense in which a property decomposes into others. By contrast, there is one clear sense in which *Plato snores* decomposes into Plato and the property of snoring. For these reasons, it would not be advisable to *identify* (#) with P.

Given this, here is what I would propose. (#) is identical with a set S *one* of whose members is P. The other members of S are Plato and the property of snoring. Remember, Plato is a property. Let Pp be this property, and let Ps be the property of snoring. If P is instantiated, then obviously Ps and Pp are also

instantiated – any state of affairs sufficient for Plato's snoring is sufficient for the occurrence of Plato and for an occurrence of snoring. So *P is instantiated* entails *Pp and Ps are instantiated.* Thus, since the instancing of P is necessary and sufficient for the truth of (#), it follows that the instancing of P *and* Ps and Pp are also thus necessary and sufficient.

P is *true* exactly if all of S's members are instantiated. So P is identical with a set S whose members are Pp, Ps, and P; and P is true in w exactly if all of these properties are instantiated.

S has exactly the right decomposition. Among its ultimate constituents are Plato and the property of snoring. Further, if all the members of S are instantiated, that is necessary and sufficient for the truth of (#). So there is, so far as I can tell, no barrier to identifying S with (#) and to identity (#)'s being true with S's being such that its members are all instantiated.

§ Now we can say why, on our analysis, it is possible to distinguish between *analytically equivalent propositions.* As we noted, this is a distinction that many theories of the proposition have a hard time accommodating.

There is obviously a difference between the proposition

(#) *that Plato snores*

and the proposition

(##) *that Plato snores and triangles have three sides.*

One of these propositions is about triangles; the other is not. Our analysis easily accounts for the distinction. (#) is a set S of the kind just described. (##) is a set S* that has *some* members in common with S, but that also has other

members. Among S*'s members are the property of being a triangle, the property of being a side, the property of being a trio (or of being thrice-instantiated). None of these things is a member of S. So and S and S* have different memberships, explaining why (#) and (##) have different decompositions.

There is more to say about the relation between (#) and (##). But it will have to wait until after we've dealt with molecular propositions. But given only what we just said, it should be clear why our analysis does, whereas the Lewisian analysis does not, account for the decompositional differences between (#) and (##). On Lewis' analysis, (#) is *identical* with some one property P such that P's being instantiated in w is necessary and sufficient for (#)'s being true in w. The trouble, of course, is that P's being instantiated in w is necessary and sufficient not only for (#) but also for (##) and for infinitely many other propositions. The difference between (#) and these other propositions is to be found, at least in part, in their decompositions; and our analysis has a place for those facts.

In light of these points, it should also be clear, at least to a point, why our analysis is able to distinguish between:

(i) *Alpha is a figure whose perimeter coincides with the class of all points equidistant from a given point in a plane*

and

(ii) *Alpha is a closed two-dimensional figure of uniform curvature.*

There is some property P such that if P's being instantiated in w is necessary and sufficient for (i) and (ii) *both* to be true in w. In fact, there is no

conceivable ripple of w's quantum that could make the one true without also making the other true. So we cannot *identify* either of those propositions with the property of having a certain kind of quantum-ripple. But since the instancing of P is necessary and sufficient for the truth of each, it is not unreasonable to give it some place in our analysis of those propositions.

Given this, here is what I propose. (i) is identical with a set Si; (ii) is identical with a set Sii. Both sets contain P. But the two sets don't have exactly the same constituency. Among the members of Si is the property of being a point; but this property is not a member of Sii. Among the members of Si is the property of having uniform curvature; this is not one of the members of Si. Each of (i) and (ii) is true exactly if all its members are simultaneously instantiated. (i) and (ii) are different propositions because, as just indicated, they have different constituencies. The truth of the one is necessary and sufficient for the truth of the other because the instancing of P is necessary and sufficient for the truth of both. The main point is that our set-theoretic analysis distinguishes between analytically equivalent but distinct propositions.

Another example might be in order. The proposition

(*) *a triangle is a triangle*

is not informative. But the following proposition is informative:

(**) *a triangle is a closed figure such that any two of its sides intersect, and such that not all three of them intersect.*

Here it might be appropriate to make a few general remarks. Again consider:

(#) *that Plato snores*

There is reasonable agreement that Plato and the property of snoring are constituents of (#). Our analysis is consistent with that. But our analysis might *seem* inconsistent with another obvious fact about (#). The two constituents just mentioned *combine* to form some larger constituent – one corresponding to the words "Plato snores" in the verbal expression of (#). Here is the trouble. We identified (#) with a set S whose members are Ps, Pp, and P. On our analysis, P is an *ultimate* constituent of S – it is no less a basic constituent of S than are Ps and Pp. When we look at (#) (or, rather, at its verbal representation), it looks as though the only *basic* constituents of that proposition are Plato and the property of snoring. To be sure, (#) does have a third constituent – one that corresponds to some kind of putting together of the words "Plato" and "snores". But, by all appearances, that *third* constituent is built out of the other two: it mustn't be *basic*. But our analysis makes it basic: P is as ultimate and indivisible a constituent of S as either of Ps or Pp. Therefore, one might conclude, our analysis is wrong.

I think that this objection to our analysis involves a failure to distinguish propositions from their verbal representations. It also involves a failure to appreciate how much we rely on sheer metaphor when we discuss propositions. Finally, that objection involves a failure to distinguish the psychological facts relating to the grasping of a proposition from the facts relating to the structure of the proposition itself.

There is no doubt that (#) is in some sense "put together out of" Plato and the property of snoring. But there is also no doubt that this putting together has not the slightest resemblance to the putting together of a car or a building. The property of snoring is an abstract object; it cannot be put next to, or on top, or

underneath anything. And given any two spatiotemporal entities, no matter how you "put them together" – no matter what kind of glue you use, no matter what kind of forklift you operate – you will not make a proposition out of them. So as for the idea that the third constituent of (#) – the "complex one" – has to result from some kind of juxtaposition of the other two – that idea is a complete non-starter. That third component will not divide into Plato and the property of snoring in anything like the way that table divides into legs and a plank.

I believe that our analysis actually makes it clear what the *actual* sense is in which Plato and the property of snoring combine to form a third constituent. If P is instantiated, that is *ipso facto* sufficient for the instantiating of Pp (i.e. Plato) and of Ps (i.e. the property of snoring). There is obviously a dependency relation among these various properties. These dependencies are given by entailments. Now properties cannot literally be parts of other properties. A property is an abstract object, and cannot be a part of anything in the sense in which a leg is a part of a body. When we talk about properties being "parts" of other properties, we must surely be talking about relations of *dependence*. These relations are often, though not always, embodied in our verbal definitions: a "triangle" is a "three-sided closed, straight-edged, planar figure". We might say that the property of being a side is a "part of" the property of being a triangle. But we are using the word "part of" metaphorically here, since abstract objects, like properties, don't have parts in any ordinary sense of the word. The *non*-metaphorical meaning here must be this: *if* x has the property of being a triangle, *then* x has the property of having straight-sides. So when we talk about properties being parts of other properties, we must, I think, be talking about relations of dependence or entailment holding among instances of those properties.

That relation does seem to hold among P, Ps, and Pp. If P is instantiated,

so *ipso facto* are the other two properties – just as when the property *triangle* is instantiated, so are *side*, *three*, *straight* and so on.

Of course, in the verbal representation of (#), the part corresponding to P is literally composed out of the parts corresponding to Ps and Pp. But that obviously doesn't mean that P itself is literally composed out of those properties. Further, that fact about language actually supports our analysis, if only weakly. The kind of dependence-relation just discussed is often verbally expressed through the sort of *verbal* juxtaposition in question: the expression "three-sided closed, planar, straight-edged figure" picks out the property of triangularity. This might make it seem as though the property picked out by that long expression literally *composed* of the properties *straight-edged*, *planar*, and so on: they are literally *parts* of the property of triangularity. But the only sense in which the properties *planar, straight-edged*, and so on, are *parts* of the property of triangularity is that some *dependence* relation holds among these properties: *if* a thing is a triangle, that entails that it is also planar and straight-edged, and so on. So the *verbal* decomposition of "Plato snores" into "Plato" and "snores" corresponds not to a relation of *part*-hood, in any literally sense, but to a relation of dependence. And that is precisely the relation that holds between P and Ps and Pp.

We are inclined to say that Ps and Pp are the only *basic* components of (#), and that any other components are "built out of" those two. The only sense that can be attached to this is: if (#) comprises any third component, then there some kind of asymmetrical relation of dependence—necessitation or entailment -- between that third component and the other two. On *some* interpretation of the expressions "less basic than" and "composed out of", the property picked out by "three-sided closed, planar, straight-edged figure" is less basic than, and is composed out of, the properties picked out by "three-sided", "closed", and so on. But all this means is that *x is triangle* entails *x is*

straight-edged (closed) but not vice versa. If we insist on understanding the relation of “composition” in more literal terms, we will inevitably fail. Obviously the aforementioned relation of dependence *does* hold between P, on the one hand, and Ps and Pp on the other. If P is instantiated, that requires that Ps (Pp) be instantiated, but not vice versa. Our analysis of (#) is consistent with that: on our analysis, the component whose verbal expression consists of *both* “Plato” and “snores” stands in exactly that relation of dependence with respect t the component whose verbal expression is “Plato” and also with respect to the component whose verbal form is “snores”. So our analysis *does* capture the only meaningful sense in which certain components of that proposition are more “complex”, or our “built out of”, others. It is true that, on our analysis, P is an ultimate constituent of (#). But that is in no way inconsistent with standing in the dependence relations just described with respect to Pp and Ps; and, as we’ve seen, to the extent that it has meaning to say that P is a “derived” constituent of (#), and not a basic one, what is meant can only be that P stands in the afore mentioned dependency relations with respect to these other constituents. Since our analysis captures that fact, our analysis captures the only meaningful sense in which certain constituents of that proposition are “built out of”, and are thus “less basic than”, others.

§ We must take care to distinguish facts relating to the *epistemology* of propositions from facts relating to the structure of propositions themselves. A failure to make this distinction is, I believe, abetted by a tendency to put too much stock in the verbal representations of propositions. When you read the *sentence:*

(#s) “Plato snores”,

you “piece together” the meaning of the whole sentence on the basis of the meanings of its parts. This, of course, might leads one to think that the meaning of the whole is put together out of the meanings of the parts in something like the way that a table is put together out of legs and a slap of wood.

When we say that you “piece together” the meaning of the whole on the basis of the meanings of the parts, the only thing this could mean is that you *figure out* the meaning of the whole sentence on the basis of the meanings of the parts. When we say that you figure this out, we can only be saying that you have hit on some dependency relation holding between the meanings of the parts and that of the whole. Once again consider the expression: “three-sided closed, planar, straight-edged figure”. That expression, as a whole, picks out the property of triangularity. On the basis of the meanings of the parts of that expression, plus a small amount of mathematical acumen, you can figure out what the expression as a whole picks out. But this obviously doesn’t that the referent of the whole expression is, in any literally sense, put together out of the meanings of the parts. You can *derive* the meaning of the whole on the basis of the meanings of the parts; and the fact that this can be done embodies the fact that there is a dependency relation between the thing picked out the by whole and the things picked out by the parts. But that is the *only* sense in which the thing picked out by the whole is “composed of” the things picked out by the parts. The *only* sense in which it is meaningful to say that one property is “composed out” of another is to say that there is some kind of dependency relation of the kind described.

Let us sum up. The fact that we *figure* out the meaning of a whole expression on the basis of the meanings of the parts means only that some (asymmetrical) dependency relation holds between the meaning of the whole and the meanings of the parts. It does *not* mean that the whole is, in any other

sense, less basic than the parts. Indeed, in this context, the only meaningful sense that can be attached to phrases like “less basic than” and “is a complex consisting of” concerns relations of dependency. In the only sense of “more basic than” that has any significance in this context, our analysis is consistent with the fact that some constituents of propositions are “more basic than” others. In the only relevant sense of “more basic than”, our analysis captures the fact that Plato (Pp) and the property of snoring (Ps) are more basic than P. Our analysis captures the fact that P’s being instantiated requires Pp’s (Ps’s) being instantiated, but not vice versa. It is true that, on our analysis, P and Ps and Pp are all equally members of S: from that viewpoint, none is more basic than any of the others; from the viewpoint, there is no stratification. S is a kind of democracy, and P is as lowly as Ps and Pp. But that doesn’t matter. The only relevant sense in which P is to be put on a higher level than Ps or Pp is that that instantiating of P instantiated necessitates that of Ps (Pp) but not vice versa. Our analysis captures that fact in spades. Every other kind of stratification or sub-ordination is irrelevant metaphor. So whatever our feelings out it might be, there is nothing logically wrong with identifying (#) with a set in which P, Pp, and Ps are all on an equal footing: their *inequality* – the asymmetry of the entailment relations holding among them -- is inseparable from them, and needn’t be buttressed by encasing them in some stratified structure.

§ These points help us deal with another important fact. Propositions are *structures;* they are *orderings*, not heaps, of constituents. But if we focus too much on language , we are likely to misunderstand the nature of this ordering.

Consider the sentence “Smith punches Jones”. The order of the words has nothing to do with the order of the corresponding entities in the proposition. In some languages, the order the nouns is reversed; in others, there are (virtually) no constraints on word-order. The sense in which Smith, Jones, and the

relation of punching are “ordered” in

(Ap) *Smith punches Jones*

has nothing to do with order in the spatio-temporal sense.

Once again, the kind of order in question has to do with relations of dependence – relations of entailment. It is true that, in (Ap), Smith has one “position” and Jones has a different one. In

(A*p) *Jones punches Smith*

Smith and Jones have switched positions. What, in non-metaphorical terms, what does this all amount to? We saw earlier that, when we are discussing propositions, facts about “composition” and “stratification” are really facts about dependence. The same thing is true here facts about the “ordinal” properties of propositions are really facts about entailment relations. (Ap) entails that Smith punches something; (A*p) does not entail this. (Ap) entails that Jones is punched by something; (A*p) does not entail this. And so on. When we say that, in (Ap), Smith is in one “position” and Jones is another, what we are saying is that, if (Ap) is to be true, then Jones must be doing the punching and Smith must be receiving it. We are talking about what is entailed by the truth of the proposition. So we are talking, not about position in any literal sense, but about relations of dependence and entailment. The ordinal properties of propositions are fixed by their inferential properties.

Consider (Ap). We pretheoretically know that Smith is in the “subject” position and Jones is in the “object” position. If we identify (Ap) with something x, then we must be able to retrieve the ordinal information that we pretheoretically have about (Ap) from x. The ordinal properties of (Ap) must

be fully “coded” in x. In general, an analysis of what propositions are must fully account for their ordinal properties.

I believe our analysis satisfies this requirement. According to us, (Ap) is identical with a set S whose members are the following properties (remember that individuals are in fact properties)

Ps: Smith
Pj: Jones
Pp: the relation of punching
P: the property that a world has if, in it, Smith is punching Jones.

And for S to be true is for all of those properties to be instantiated. Of course, if all of those properties are instantiated, then P is instantiated. If P is instantiated, there are various consequences: Smith is punching something, Jones is being punched by something, and both punchings are one and the same. We saw a moment ago that the ordinal properties of (Ap) lie in these consequences. So by virtue of identifying (Ap) with S, and by identifying the truth of (Ap) with the instantiating of all of S’s members, we have ensured that S codes all the ordinal information characteristic of (Ap).

§ Here it might be appropriate to point out what are, I believe, some virtues of our analysis. Remember that theory (i), and also Wittgenstein’s “picture theory, couldn’t account for the existence of false propositions, since they identify propositions with their truth-makers. On our account, the *truth* of a proposition requires the existence of some spatio-temporal state of affairs. But the proposition *itself* requires only the existence of properties: it requires only the existence of *ways* that mass-energy might be displaced.

Molecular Propositions

Let us discuss *molecular* propositions: these are propositions that are formed out of other propositions, e.g. *Bob is tall and Sam is short*. In keeping with tradition, I will describe quantified propositions (e.g. *something is short*) as "molecular".

A point about methodology is in order. When dealing with atomic propositions, we defined truth in terms of instantiatedness: a proposition is a set of properties, and that proposition is true iff all its members are instantiated. We must treat molecular propositions in the same way. We cannot define *proposition* and *truth* one way for atomic propositions and another way for molecular propositions. That would be absurd. Propositions are propositions. Truth is truth. But this need for uniformity requires that, initially, we say a few counterintuitive things.

Consider the proposition:

(jsw) *That John believes that snow is white.*

This is a proposition about a proposition; it attributes a property – being believed by John – to *that snow is white.* Obviously that property and that proposition are constituents of jsw. So jsw is identical with a set S among whose members are that property and that proposition.

Here a refinement is in order. We have identified propositions with sets of properties: it is crucial to our analysis that every component of a proposition be something that can be instantiated. But, as we've analyzed it, the proposition *that snow is white* is not exactly a property; it is a set. (I don't think that properties *are* sets.) In order to provide a uniform analysis, I propose this. What is a constituent of jsw is the property of being identical with *that*

snow is white. That property is instantiated exactly if that proposition *exists.* We must be careful here: the property of being identical with *that snow is white* can be instantiated without that proposition's being *true.* For the instancing of that property, it is necessary and sufficient only that *that snow is white* exist. To sum up, jsw is a set S; and among the members S are the property of being believed by John and also the property of being identical with *that snow is white.*

But those cannot be the only members of S. Both of those properties can be instantiated *without* John's believing that snow is white. The property of being identical with *that snow is white* is automatically instantiated. If John believes that grass is green (or even that snow *isn't* white), then the property of John's believing something is instantiated. So the properties in question can be instantiated with jsw being true. S must comprise some other constituent. This other constituent is not hard to identify.

In worlds where John believes that snow is white, the quantum is rippled one way; in other worlds, the quantum is rippled differently. Let P be the property that a world has if, in it, the quantum is rippled in the way just mentioned. So P is instantiated in w exactly if, in w, John believes that snow is white. Among the members of S are P. Of course, if P is instantiated, then so are the properties that are members of S. Obviously if P is instantiated, then John believes something or other (so the property of being believed by John is instantiated). And the property of being identical with the proposition *that snow is white* is instantiated under any circumstances. Thus if P is instantiated, so are the other members of S. And jsw is true exactly if P is instantiated. So jsw is true exactly if all three members of S are instantiated. Thus we identify jsw with S, and we identify S's being true with all its members being instantiated.

The question arises: Why not just say that P is the *sole* member of S, since

P’s being instantiated is necessary and sufficient for the truth of S? We’ve already seen the answer to this. It is pretty clear that jsw comprises a distinct constituent corresponding to the property of being believed by John, and it also comprises a distinct constituent corresponding to the proposition *that snow is white* (and, thus, to the property of being identical with that proposition).If we identify jsw with a set, then we need the membership of that set to clearly reflect these facts about jsw’s constituency. If that set’s sole member is P, then those facts are eclipsed; the set doesn’t have the right decomposition.

A given propositions is “composed” of various things. We saw earlier that this composition-relation is best understood in terms of relations of dependency or entailment holding among those constituents. It is pretty clear that jsw comprises one constituent corresponding to the words “john believes”, another constituent corresponding to the words “that snow is white”. It is *also* clear that jsw comprises some more “complex” constituent corresponding to the juxtaposition of “John believes” and “that snow is white”. If that more complex constituent exists, then so does each of the simpler ones; but both of the simpler ones can exist without the complex one existing. The thing corresponding to “John believes” can exist simultaneously with the thing corresponding to “that snow is white”. But that doesn’t guarantee the existence of the thing corresponding to the juxtaposition of those expressions. So an asymmetrical relation of dependency holds between that more complex constituent and those less complex
ones.

Exactly these points apply to the members of S. If P is instantiated, then so is each of the two properties just mentioned. But both of those other properties can be instantiated with P being instantiated. So there is a dependency relation between P and each of the other two properties just mentioned; and that relation perfectly parallels the one discussed a moment ago – both relations

are asymmetrical in precisely the same way.

Given this, it becomes clear why we must identify jsw with S. S comprises three constituents: one corresponding to the proposition *that snow is white*; another to the property of being believed by John; and a third corresponding to something that, in the relevant sense, combines the two. (Remember that the kind of combining in question consists in the holding of dependency relations: and the right kinds of dependency relation holds among the three things just discussed.) So S has the same constituency as jsw. Further, jsw is true exactly if all three members of S are instantiated. So we can identify jsw's being true with its being the case that all the members of S are instantiated – roughly, we identify truth with instantiatedness. And we identify jsw with S. In this way, we do justice to the decompositional fine-grain of jsw, and also to its truth-conditions.

At this point, I must address what many will find to be a worrisome feature of our analysis. We have identified propositions with sets. Thus the proposition:

(jsw) *John believes that snow is white.*

is a set. We decided that what is a member of that set is not, strictly speaking, the proposition *that snow is white*, but rather the *property* of being *identical with that proposition.* Obviously this proposal is counter-intuitive. One would think that the proposition iself ought to be a member of that set.

There are a couple of things to say here. First, on one delineation of the expression "constituent", we are making the proposition *that snow is white* be a constituent of (jsw). But, even if we leave that aside, I think that the just mentioned counter-intuitiveness of our analysis is simply another embodiment of some very crude and false theories that we tend to have about the sense in

which things are “constituents” of propositions. *That Plato snores* is not built out of Plato the way a building is made out of bricks. (jsw) is not built out of *that snow is white* the way my car is built out of metal parts. As we saw earlier, whenever we relate the notions of “constituency”, “composition”, and “part” to propositions, we are really talking about relations of logical dependence and entailment. On our analysis, the truth of (jsw) absolutely *depends* on some fact about the proposition *that snow is white.* On our analysis, (jsw) is true exactly if *that snow is white* has the property of being believed by John. In so far as our analysis of (jsw) is consistent with that fact, it fully accommodates the fact that *that snow is white* is “constituent” of (jsw) – for that “constituency” relation just *is* the dependency relation discussed a moment ago. (In any case, that is one way to delineate the concept of “constituency”, as it relates to propositions.) Now *in order* to accommodate that fact, we had to identify jsw with a set S that comprises, not *that snow is white*, but instead the property of being identical with that proposition. But this is in no way inconsistent with the truth expressed by statements like “the proposition that snow is white is a constituent of jsw”: for on any interpretation in which such a statement is meaningful, let alone true, that statement amounts to an affirmation of the dependence relation described a moment ago; and, as we just saw, our analysis fully accommodates the existence of that dependence relation. So by identifying jsw with a set S one of whose constituents is the property of being identical with *that now is white*, our analysis actually *captures* the truth of the statement that *that snow is white* is a constituent of jsw, in so far as that statement is meaningful and true.

What we said about jsw applies, with only a few obvious changes, to all molecular propositions. Consider:

(sn) Sam is not short.

This attributes falsity to the proposition *that Sam is short.* So among sn's constituents are that property and that proposition. For reasons discussed earlier, in connection with jsw, it would be preferable to say that the property of being identical with *that Sam is short* is a constituent of that proposition. So sn is a set S that has as members (inter alia) the property of being false and the property of being identical with the proposition *that snow is white.* (The property of falsity is instantiated in a world w iff some proposition is false in w. Presumably that property is instantiated in every world. But that needn't concern us here.) Like any other proposition, sn is true iff all the members of S are instantiated. But obviously sn could be false even if both the property of being false and the property of being identical with *that Sam is short* were instantiated. (After all, those properties are instantiated in w if it is false in w that Smith is bald and true that Sam is short.) So S must comprise some third constituent.

In light of our earlier points, this third constituent is not hard to identify. Consider the way the quantum is rippled in a world in which Sam is not short, i.e. in which Sam is of average or above average height. Let P be the property that all and only such worlds have. P's being instantiated in w is necessary and sufficient for sn's being true in w. So P must be a member of S. But we don't want to make P be the only member of S. For now familiar reasons, making P be the sole member of S wouldn't do justice to the decomposition of sn. But we are being true to that decomposition if we make P *and* the other two properties (that of being false and that of being identical with *that Sam is short*) be the membership of S. So has the same decomposition as sn. Further, sn is true exactly if all of those three properties are instantiated. So we identify S with sn, and identity sn's being true with S's being such that all its members are instantiated.

§ Before dealing with quantified propositions, let us deal with conjunctions. Consider:

(sb) *That Sam is tall and Bob is short.*

First of all, this proposition is true exactly if the two constituent propositions are true. So the constituent whose verbal representation is “and” can be treated as a property that ascribes truth to a pair of propositions exactly if both members of that pair are true. This analysis is, I will argue later, equivalent to the familiar idea that “and” denotes a function that assigns truth to a pair of propositions exactly if both members of that pair are true.) Let AND be that property. The property AND is instantiated in a world w exactly if some pair of proposition is true in w. (Presumably AND is instantiated in every world. But this needn’t concern us.)

Obviously (sb) has AND as a constituent. For now familiar reasons, (sb) also comprises the property of being identical with *that Sam is tall* and also the property of being identical with *that Bob is short.* Let Pst and Pbs be these two properties. So, for now familiar reasons, sb is identical with a set S whose membership includes AND and Pst and Pbs. But – once again for familiar reasons – those cannot be the only members of S. After all, all of those properties could be instantiated with sb’s being true. So S must contain a fourth constituent. Consider the way that quantum is rippled in a world where it is the case both that Sam is tall and Bob is short. Let P be the property had by all and only such worlds. sb is true in a world w exactly if P is instantiated in w. (Of course, is P is instantiated in w, then so are AND and Pst and Pbs.) If we suppose that S comprises P, then sb is true exactly if every member of S is instantiated. Further, S has the same decomposition as sb – both comprise

distinct and separable parts corresponding to the property picked out (in English) by “and”, and also to the proposition *that Sam is tall* and the proposition *that Bob is short.* So we might as well identify sb with S, and identify the truth of sb with its being the case that every member of S is instantiated.

Similar discussions provide the analysis of propositions like *Bob is short or Sam is tall, Bob is short because Sam is tall, Bob is short if Sam is tall*, and so on. There are no conceptual obstacles here, though some care must be taken in choosing the constituents corresponding to the connectives.

§ Quantified Propositions

In keeping with tradition, I regard quantified propositions as being molecular. The proposition *nothing snores* is in the same category as *Bob snored and Fred snores*, and not in the same category as *Bob snores.* I adopt this view partly because, as we will see, it allows for a smooth and uniform semantics. (So the reasons are somewhat like the reasons for seeing n^0 as being equal to 1, for any natural numbers n). But I also think that, at bottom, quantified propositions really are in the same category as conjunctions, disjunctions, and so forth. They are meta-propositions: propositions about classes or sequences of propositions. (I hope that reasons for this become clear.)

Let us begin our treatment of quantified proposition with:

(ns) *that nothing snores.*

I think it is important to begin with negative existentials, since they make obvious certain features of quantified propositions hidden by existentials and

universals. Following Frege, I see (ns) as saying that every proposition having the form *x snores* is false. For now, I would like permission to treat the form *x snores* (sometimes referred to as a "propositional function" or "open proposition") as an entity in its own right. In a moment, we will discuss what kind of entity it is.

On the face of it, (ns) has (inter alia) two constituents: the propositional form just mentioned, and the property of being universally false, i.e. the property had by a thing x if every instance of x is false. So the form *x is a square circle* has that property, since every instance of that form is false. Let us refer to that property as UF. UF is instantiated in a world w iff there is some propositional form such that, in w, every instance of that form is false. (Presumably UF is instantiated in every possible world. But, once again, this needn't concern us.) For now familiar reasons, we must identify ns with a set S whose membership includes (inter alia) UF and the form *x snores.*

There is a caveat. We have said that a proposition is a set of propositions, and is true exactly if all those properties are instantiated. Here is the problem. Let NS be the set with which (ns) is identical. We don't want to say that the property of snoring is a member of that set. Why not? Because if that property is instantiated, then (ns) is automatically false. So we don't want to say (ns) has that property as a member.

It is not hard to deal with this. What is a member of S is not exactly the form *x snores*, but rather the property of being *identical* with that form.

This procedure is exactly analogous to one we used earlier. We decided that (stbs) *that Sam is tall and Bob is short*
is identical with some set S; and decided that what is a member of S is not exactly the *that Sam is tall*, but rather the *property* of being identical with *that Sam is short.* (Let Pss be that property.)

We saw that, contrary to first appearances, this procedure actually

accommodates the fact that *that Sam is tall* is a "constituent" of (stbs), at least on one viable delineation of the term "constituent". For that procedure *accommodate* the dependence relation that holds between (stbs) and *that Sam is tall*; and, as we saw, that dependence-relation is really identical with the constituency-relation in question. Further our decision to make Pss be a constituent of S enabled us to extend our analysis of truth as instantiatedness to molecular propositions. That decision enables us to capture both the decomposition and the truth-conditions of (stbs): and those are the two important desiderata.

Let us return to

(ns) *nothing snores.*

We want our analysis of (ns) to do justice both to its decomposition and to its truth-conditions. This can be done, provided we apply the procedure (mutatis mutandis) just described. We identify ns with a set S that comprises (inter alia) UF and also the property of being identical with the form *x snores.*

Before we continue, a somewhat delicate point is in order. As we will see, propositional forms probably *are* properties. I believe that *x snores* is the property had in common by all and only *that John snores, that Mary snores, that Frank snores*, and so on. I believe that an *instance* of that form is some *specific* proposition of that form. So an *instance* of *x snores* is a constituent of *that John snores.*

But *that John snores* is not an *instance* of the property of being identical with *x snores*. An instance of the *latter* property is the form *x snores* itself.

An example may clarify my meaning. An instance of the property *even number* is the number four. But the number four is not an instance of the property *identical with the property of being an even number.* After all, the number four is not *identical* with the property of being an even number.

Similarly, *that John snores* is not an instance of the property *identical with the propositional form: x snores*, even though it is identical with an instance of the form *x snores.*

To sum what we've said thus far: ns is identical with a set whose membership includes the property of being identical with the form *x snores* (let Pxs be that property) and also includes UF. I propose that, by doing this, we can account for the truth-conditions and the decomposition of ns.

But, again for familiar reasons, S cannot comprise *only* UF and Pxs. After all, *those* two properties could be instantiated in a world w where ns was false. Suppose that, in w, the proposition *nothing plays tennis* is true. In that case, UF is instantiated in w. Further, suppose that Mary snores in w. In that case, Pxs is instantiated in w. (Of course, Pxs can be instantiated in a world even if *nothing* in that world snores. For, it seems to me, even in a world where nothing snores, the property of being *identical* with the form *x snores* exists. But, once again, we needn't dwell on this.) So both UF and Pxs can be instantiated in a world without ns being true. So ns cannot be identical with a set that contains *only* UF and Pxs.

So given that we are identifying ns with a set, that set must contain some third constituent. Again, this is not hard to identify. Consider the way the quantum is rippled in a world where something snores. Let P* be the property that all and only such worlds have. Now let P be the property that all and only worlds lacking P* have. So a world has P iff its quantum is appropriately *un*rippled. ns is thus identical with a set S that comprises UF and Pxs *and* P. ns is true exactly if P is instantiated. Of course, if P is instantiated, so *a fortiori* are UF and Pxs. It follows that ns is true exactly if P and Pxs and UF are instantiated. But, for now familiar reasons, we don't want to say that S's sole member is P: that wouldn't do justice to the compositional structure of S. So we identify ns with a set S containing all three of the properties just

mentioned; and we identify ns's being true with S's being such that all its members are instantiated.

§ Our analysis made heavy use of the idea that there is such a thing as the propositional *form* x snores? Now we must answer the question: what is that thing?

Consider the class of propositions containing all and only *that John snores, that Mary snores*, and so on. The form *x snores* is the property had in common by all and only those propositions. In light of what we said earlier, we can give a precise statement of what this property is. Remember that the proposition *John snores* is identical with a set S whose members are: the property of being identical with John; the property of snoring; and the property P had by world w exactly if, in that world, John snores. An exactly similar analysis applies to *that Mary snores* and *that Ethel snores* and so on. So the property of being identical with the form *x snores* is one and the same with the property of being a set S such that, for some individual x, S contains (i) the property of being identical with x; (ii) the property of snoring; and (iii) the property P had a by world w exactly if, in that world, x snores.

§ The proposition:

(es) Everything snores

is dealt with in a way exactly analogous to the way we dealt with (ns). First of all, (es) is true if the propositional form *x snores* is "universally true", i.e. is such that each instance of that form is true. Let UT be the property of universal truth. We identify (es) with a set S that contains UT and, for reasons already seen, that also contains Pxs (the property of being identical with the property *x*

snores). For now familiar reasons, S must contain a third constituent. Consider the way the quantum is rippled in a world w in which everything snores. Let P be that property. S contains P and UT and Pxs. (es) is true exactly if P is instantiated. If P is instantiated, then so are Pxs and UT. So (es) is true exactly if all of P and Pxs and UT are instantiated. Thus (es) is true exactly if all the members of S are true. Thus we identify (es) with S, and identify the truth of (es) with its being the case that all the members of S are instantiated.
§ Let us now consider the proposition:

(ss) *something snores.*

Here it is tempting to say this:

It is obvious what that proposition is. It is identical with the property of snoring. And that proposition is true exactly if that property is instantiated. So (ss) fits perfectly into our view that propositions as sets of properties and that for a proposition to be true is for the corresponding properties to be instantiated.

But such a view is not correct. Presumably *something snores* has the same basic structure as *nothing snores* and *everything snores.* For reasons we have seen, the latter have as members, not the property of snoring, but rather the property of being identical with that property. We want (ss) to have a structure analogous that of those other two propositions.

Also, (ss) has a certain amount of decompositional structure. The property of snoring doesn't have any decompositional structure, at least not in the sense in which (ss) does.

Given these points, this is what we say. (ss) is identical with a set (SS). (SS)

has the property of being identical with the property of snoring as a member. (SS) also has the property of being instantiated as a member. Let (PT) be this property.

(ss) also has a third member. Consider the way the quantum is rippled in a world where something snores. Now consider the property common to all and only worlds having that property. Let (s*) be that property. For exactly analogues of reasons considered earlier, in connection with (ns) and (es), (SS) has (s*) as a member.

So the proposition *something snores* is identical with a set that has the following three things as members: the property of being identical with the property of snoring; (s*); and (PT). That proposition is true in a world exactly if each of those properties is instantiated in that world.

This analysis is, admittedly, counter-intuitive. But it is demonstrably consistent with both the truth-conditions had by *something snores* and also with facts about its decomposition.

Here one consequence of our analysis should be made explicit. According to our analysis, the proposition *that John snores* is identical with a set that has the property of snoring as a member. (And our analysis also requires that that property be instantiated if that proposition is correct.) But on our analysis, (ss) is identical with a set that does *not* have that property as a member. And this will strike many as odd, and even unacceptable.

Let us recall a point made earlier. The *John snores and Fred snores* is *obviously* built out of other propositions. It is *obviously* "molecular". Traditionally, semanticists have regarded quantified propositions (like *something snores*) as also being molecular, even though they are not, strictly speaking, built out of other propositions. I think that this tradition embodies a real insight into the nature of quantified generalization. When you say "something snores", you are, arguably, making a statement about a *class* of

propositions: the class of propositions of the form *x snores.* You are saying that at least one such proposition is true. So quantifications are meta-propositions. For reasons given earlier, the same is true of conjunctions, conditionals, negations, and the like. The proposition *if grass contains chlorophyll*, *then grass is green* affirms the existence of a relation between propositions. It is thus a meta-propositional proposition. The right definition of "molecular proposition", then, is probably not "proposition built out of other propositions", but rather "proposition that concerns propositions". As we just saw, quantified generalizations fall into this category, as do our paradigms of molecular propositions (conjunctions, conditionals, disjunctions, and so on). So the tradition of assimilating those two classes into a single proves correct or at least theoretically very convenient.

§ Now we can give fairly precise answers to two questions. First, what is it for E to "denote" O? Second, what is the semantic function of *grammar*?

There are two delineations of the term "denote". One of these, E denotes O exactly if, in virtue of having the form "…E…", a sentence-token encodes a proposition P such that P's being true involves the existence of a state of affairs such that O is a constituent of that state of affairs. This is the sense in which "Smith" denotes Smith. Consider the sentence "Smith punches Jones." Suppose that (Ap) – the proposition encoded in a token of that sentence – is true. P is true in virtue of the fact that there is a state of affairs that (inter alia) has Smith as a component.

Let us refer to this as "denoting1". What does the expression "red" denote1? The obvious, but wrong, answer is: "the property of redness". The right answer is this. A token of "red" denotes1 an *instance* of that property.

Consider a token of the sentence “that car is red”. If true, that token encodes a proposition P such that P holds in virtue of the existence of a state of affairs of which some constituent is an instance of redness.

To understand the other delineation of “denote”, we must remember that each proposition is identical with some *set.* On this delineation, E denotes O exactly if, in virtue of having the form “…E…”, a sentence-token encodes a proposition P such that the set with which P is identical has O as a member. Let us refer to this relation as “denoting2”. Consider a toke of “Smith is tall”. The occurrence of “tall” denotes2 the *property* of tallness, not some instance thereof. And the occurrence of “Smith” denotes2 the property of being identical with Smith, not Smith himself (i.e. not some instance thereof). Here is the general rule.

E denotes1 O exactly if, in virtue of having the form “…E…”, a sentence-token encodes a proposition P such that, if P is true, it is true in virtue of the existence of a state of affairs one of whose constituents is O.

E denotes2 O exactly if, in virtue of having the form “…E…”, a sentence-token encodes a proposition P such that O is one of the members of the set with which P is identical.

§ What do grammatical inflections do? “Smith punches Jones” encodes a proposition; “Smith, the relation of punching, Jones” does not. The first has the right grammatical inflections; the second does not. The difference between a sentence-token and an unmeaning heap apparently lies in facts about grammar. Grammatical facts code the relevant ordinal information.

We’ve seen what that information is.

(Ap) The proposition *Smith punches Jones.*

Ps: the property of being identical with Smith.
Pj: the property of being identical with Jones.
Psx: the property of being a thing x such that Smith punches x.
Pxj: property of being a thing x such that x punches Jones.
Pp: the relation of punching.
P: as previously defined.

(Ap) is a set S whose members are exactly the properties just listed, and for (Ap) to be *true* is simply for all of its members to be instantiated.

Let t* be a tokne of "Jones punches Smith", and let (Bp) be the propositoni encoded therein. (Bp) and (Ap) will have different constituents. t and t* differ in the order in the component expressions are arranged. This is a grammatical difference; it is the way in which, in English, facts about *case* are expressed. In other languages – Latin, Russian, Greek – those facts are coded primarily in word-endings, not (so much) in facts about word-order. In English, word-order does exactly what is done, in languages ike Russian and Latin, by patently grammatical inflections. So word-order is a grammatical device *par excellence.*

What does word-order (in English) do? The set with (Bp) is identical will have different members from the set with which (Ap) is identical. For example, (Ap) has as a member:

Psx: the property of being a thing x such that Smith punches x. property of being a thing x such that Smith punches x.

(Bp) does not have Psx as a member. (Bp) has as a member:

Pjx: the property of being at thing x such that Jones punches x.

(Ap) does not have Pjx as a member:

Grammatical inflections, then, do have an extremely significant semantic role: they determine what will be the constituents (on any delineation of that term) of the propositonis meant by sentence-tokens. In t, the facts about word-order tell you (inter alia) that Psx is a member of the corresponding proposition. In t*, facts about about word-order tell you (inter alia) that Pjx is a member of the corresponding proposition.

Are grammatical inflections referring terms? There are two ways we can go about answering this question.

Grammatical inflections do not *by themselves* refer to anything; but they *assign* referents to sequences of expressions on the basis of the referents of those expressions. Consdier t again. As a whole, t denotes the proposition (Ap). The grammar of t – the word-order, the inflection on the verb – do not themselves denote anything. But, on the basis of the referents of “Smith”, “love”, and so on, they assign referents to various complex expressions. For example, the proposition (Ap) is assigned to the whole sequence. The property Psx is assigned to the complex expression “Smith punches”, on the basis of the referents of “Smith” and the root-verb “punch”. The property Pxj is assigned to “punches Jones” on the basis of the referents of “Jones” and the root-verb “punch”.

Very roughtly, the general idea would be this. Grammatical inflections do not themselves refer. But when a grammatical inflection occurs in a sequence of referring terms, that occurrence assigns a referent to that sequence as a whole on the basis of the referents of the expressions composing that sequence.

There is another way to look at facts about grammar. The idea would be the grammatical inflections *do* refer; but what they refer to is a function of the verbal environment. Consider the inflection on the verb in t. We have seen that, because that inflection occurs in that place, the corresponding proposition has Psx and Pxj as members. So we *could* conceivably see that inflection as denoting (or, strictly, as denoting2) those properties. When that same inflection occurs in a different environtment, e.g. a token of "Plato snores", it denotes a different thing.

I that the first analysis is the preferable one: it is, I think, more natural to see grammatical inflections, and the like, *not* as themselves referring to anyting, but as assigning referents to sequences of expressions on the basis of the referents of the components of those sequences.

In effect, we've already seen why this path is preferable to the other. If we taken other other path, then we must say that an occurrence of a grammatical inflection denotes a number of different things – e.g. the ending on the verb in t denotes both Psx and Pxj. That inflection is surely not *ambiguous* between those two properties. At the same time, we probably don't want to say that it denotes some *complex* entity consisting of both those properties. What would that entity be? Would it be some set consisting of exactly of those properties? If we agreed to such a proposal, that would unnecessarily complicate our analysis of propositions. So it seems better to say that, strictly speaking, grammatical inflections do not themselves refer, but that they *assign* referents to sequences of expressions on the basis of the referents of the components of those sequences. At the same time, the facts do not *compel* us to say this: such inflections *can* be treated as contextual referring terms – terms whose referents are a function of the verbal environment in which they occur. The obstacles to taking this path are considerations of simplicity and the like – so far as I can tell, it would not be *wrong* to do so. But for the reasons given a moment ago,

my official position is that it is best, from the viewpoint of theory-construction, not to regard grammatical inflections as referring terms.

Some Metaphysical Consequences

Thought separates things that, in reality, are not separable. We think that Smith is a discrete, isolable constituent of the universe. But he is not. Wherever Smith exists, so does a state of affairs. This is the case for a number of reasons. First, nothing is propertyless. If Smith exists, he must have a certain height, weight, and so on. So what we find in reality is never just *Smith*, but Smith's having such and such properties. This is equivalent to saying: what we find is not *Smith*, but always some state of affairs involving Smith. Also, Smith's existence supervenes on the occurrence of innumerable states of affairs: sub-atomic, atomic, molecular, cellular, metabolic, psychological. For modal reasons, Smith cannot be *identified* with the sequence of events on whose existence his own supervenes. But he is plainly inseparable from them: there but for the grace of those lower level existents goes he.

The truth is that Smith is more in the nature of an abstract. He is no separable from the existence of concrete states of affairs than are instances of whiteness or roundness. People can no more exist outside of states of affairs than smiles can exist outside of people (or Cheshire cats).

This suggests that Smith is really a *property* of states of affairs. What is it for Smith to exist in a world. The old picture is that Smith is at the bottom, and properties are heaped on top of him. This picture is reinforced by language, which gives Smith his own, semantically simple expression; and which represents the attribution of properties to Smith as the adding of gratuities to an already self-supporting entity. Presumably this fact about language has its

basis in deep cognitive facts. Be all of this as it may, it is the wrong metaphysical model. We find states of affairs first. The basic constituents of a world are displacements of mass-energy, wrinkles in space-time. If those wrinkles are of such and such a type – if those wrinkles have the right properties – then Smith exists in that world; if not, not. So Smith, I would suggest, is a property of states of affairs.

On similar grounds, I would argue that the things we call "properties" – redness, roundness, and the like – are really properties of properties. Actually, they are, I belive, properties of properties of properties -- 3rd order properties!

First of all, you never find instances of redness in isolation of states of affairs. The same is true of instances of roundness, sweetness, and anything else that we refer to as a "property". What you find is a red (or sweet or round) state of affairs.

What is a state of affairs? Pick some state of affairs. (To facilitate things, choose one that you can see or feel.) Consider the fact that there is a tree in such and such place at such and such time. Let S be that state of affairs. What is it for a world to comprise S? S's existence in a world supervenes on the existence of innumerable displacements of mass-energy – ultimately, on the presence of certain "wrinkles in the quantum" of that world. So that state of affairs exists in a world w if some of the wrinkles in that world's quantum are of the right type, i.e. if those wrinkles have the right properties. So a state of affairs is a property of a property. Instances of redness, tallness, and the like – the things we are used to describing as "properties" – are properties of states of affairs. So they are properties of of properties of properties.

Propositions as sets of properties, truth as set-member-instantiatedness (the short version)

By a proposition I mean the truth-evaluable entity meant by a sentence-token – by, e.g. an utterance of "that man is a scoundrel" or "it's hot in here right now."[384]

It is fiercely debated what propositions are. Some say that they are sets of worlds or functions from worlds to truth-values; others that they are facts; others that they are possible (but not necessarily actual) facts; others that they are sets of their own consequences; others that they are structures that contain the same parts as the facts that, if they were true, would make them so, but in which those parts are arranged very differently from the way in which they would be arranged in the corresponding facts. Finally, some say they don't exist.

My purpose here is to propose a positive answer to the question "what are propositions?" Here is that answer: Propositions are sets of properties, a proposition is true just in case the properties that are members of it are instantiated; and a corollary is that truth is identical with instantiatedness (or, more accurately, that truth is the property of being a set containing only instantiated properties). Aside from a few brief remarks, I will leave it open whether *all* sets of properties qualify as propositions. But a case will be made that *only* such sets are propositions.

A few words will be said as to why the alternative theories are unsatisfactory; but my remarks on that matter will be brief, since the deficiencies of the existing alternatives have been thoroughly discussed elsewhere.

The argument presented here will not be airtight. But, I hope, it has

sufficient *prima facie* plausibility that it is worth investigating, despite its not being cogent in every respect.

Here is an outline of an argument for our analysis. Given any proposition P, there is some set S of properties such that P is true if and only if all the members of S are instantiated. The "if and only if" is to be taken to have the strongest possible force – to denote conceptual or, in any case, metaphysical necessity. This suggests that a proposition *is* a set of properties and that a proposition's truth *is* the joint instantiatedness of its members. Hence my proposal: a proposition is a set of properties and a proposition's being true is identical with its members' being instantiated.

This analysis, I will argue, holds for all propositions; but special measures must be taken to deal properly with molecular, especially quantified, propositions.

Our analysis will enable us to see what, exactly, is meant by the obscure metaphors that dominate semantic discourse – metaphors embodied in terms such "saturate", "synthesize", and "constituent."

Our analysis is intended to do justice to the fact that propositions have decompositional structure. The proposition *Harry owns a dog* has a certain structure corresponding – in some way, to some degree – with the structure(s) of the sentences that express it. Indeed, there much be such correspondence, else people couldn't assign that proposition to those sentences and therefore couldn't understand them. Not all analyses of propositions, as we will see, do justice to the humdrum facts that sentences can be understood and that propositional structure must therefore have some kind of systematic relation to sentential structure.

Section I. A Dogmatic Statement of the Basics

In this section I will list the conceits that underlie the present analysis. Argumentation for these conceits will be given in Section II and will be quite absent from this section. There may be, and probably are, some redundancies on the list. (Eliminating the redundancies would decrease clarity less than it would increase cogency or expository elegance.)

Conceits #2 and #7 will be initially seem quite counterintuitive. But the *raison d'être* for them will become apparent in Section II.

I realize that this method of presentation is unusual. But I feel that, ultimately, it is for the best that, as soon as possible and with as little obfuscation, I lay bare the machinery driving my analysis (so that, if nothing else, my biases are exposed at the outset).

Conceit #1. A proposition is a set.

Conceit #2. Every member of such a set is a *property*. In the proposition *Plato snores*, what occurs is *the property of being identical with Plato*, not *Plato* per se. (For a proposition containing that property to be *true*, the property of being identical with Plato must be instantiated. For, for example, for "Plato is happy" to be *true*, the property of being identical with Plato must be instantiated – Plato must exist. But Plato *per se* is not a constituent of the proposition itself.)[385]

Conceit #3. A proposition is true iff each of the properties in the set identical with that proposition is instantiated. A proposition is a set and a proposition's being true is that sets' members' being instantiatedness.

Conceit #4. Each "constituent" of a proposition is either identical with or in some obvious way corresponds to a member of the set with which that

proposition is identical. Give or take some nuances, propositional-constituency is set-membership.

Conceit #5. The logical properties of a proposition are perspicuously represented by facts about the membership of the set with which that proposition is identical. The proposition *Plato snores* entails *something snores* and also *Plato does something* and, finally, *something does something.* Among the members of the set with which *Plato snores* is identical are *the property of Plato's doing something* and *the property of something's snoring* as well as *the property of something's doing something.*

What the proposition does in the way of entailing the set does in the way of instantiating.

Conceit #6. Facts about the structures of the decompositions of propositions are readily and transparently equated with facts about the memberships of the sets with which they are identical. So the proposition *Aristotle snores* must have a discrete, easily identified component in common with *Plato snores* and with *Bill Clinton snores.* ***[386]***

Conceit #7. Let p be a proposition that is a proper constituent of another proposition P. (So if P is "if Smith comes home early, then Jerry will be mad", then p might be "Jerry will be mad.") Let s be the set that is identical with p, and let S be the set that is identical with P. In that case, what occurs as a member of S is not s but is rather *the property of being identical with s.* [387]

Conceit #8. It has long been known that the "constituents" of propositions sometimes "synthesize" with one another to form new, more "complex"

constituents (these are all terms whose meanings have been left opaque).

I would like to make a proposal as to the nature of that synthesis. Consider the sentence "Plato snores." According to the analysis to be put forth, among the constituents of this proposition are "the property of being identical with Plato" and "the property of snoring." Now consider the property of being identical with Plato and also consider the property of snoring. To "synthesize" those two things –or, to use Frege's term, for Plato to "saturate" the property of snoring[388] – is to identify a property p_3 each one of whose instances is an instance *both* of snoring *and* of being Plato. In general, the synthesizing of propositional constituents consists of two things: first, realizing that all such constituents are properties; second, supposing that p_1 and p_2 are the two properties that you wish to synthesize, identifying some third property p_3 each of whose instances are instances both of p_1 and p_2. (Synthesizing new properties out of old ones doesn't consist in creating new properties, but only in identifying properties whose properties combine those of the old ones.)

Conceit #9. grammatical inflections encode ordinal information. Given a series of properties $p_1 \ldots p_n$, the grammatical inflections on the corresponding terms will say, for which j and k, to look for a property P such that each instance of P is an instance of both p_j and p_k; and those grammatical inflections will also specify when, relative to other operations of the kind just described, a given one is to occur.

Conceit #10. As a rule, discrete parts of sentences must correspond to discrete parts of propositions and, therefore, to discrete members of the sets with which propositions are identical. It must be possible to puts facts about

the structures of propositions into a transparent relation with facts about sentences.

Conceit #11: Existing propositions cannot contain non-existent parts (e.g. Socrates, Plato), this being a special case of the general principle that philosophical analyses cannot defy the laws of logic.

Conceit #12: “Plato” refers to Plato; “Plato” does not refer to the property of being identical with Plato. Contrary to first appearances, this is consistent with my analysis.. For “Plato” to refer to Plato is for the property of being identical with Plato to be a member of the set identical with the proposition meant by ⌜…Plato…⌝ and it is therefore for Plato *per se*, as opposed to property of being identical with Plato, to be a constituent of the fact that makes that sentence true (if it is true).

Section II. The Arguments

Now I will try to provide demonstrative support for the statements just made.

Consider the proposition:

1. Tommy is smoking.

Intuitively, this proposition seems to consist of at least three components: Tommy (or, perhaps, the property or concept of being Tommy), the property of smoking, and some kind of “synthesis” of the two – some complex

constituent that involves Tommy's smoking. (We will try to state precisely what this synthesis involves.)

If 1 is true, then

A. the property of being identical with Tommy

is instantiated, and so is

B. the property of smoking.

But, of course, those two properties could be instantiated in a world where Tommy is not smoking – in a world where, for example, Larry is smoking and Tommy is playing golf (and also not smoking). We thus need a third property – a property such that, if a set S contains that property along with A and B, then 1 will be true just in case all of S's members are instantiated. That third property is not hard to identify. It is:

C. The property of being a person who is smoking and is identical with Tommy.

Before moving on, we must address some possible unease that C may awaken. Few will doubt that the property of being identical with Tommy is a perfectly acceptable one.[389] But some might resist signing off on C. We mustn't be dismayed, however; for we must remember that Tommy cannot *just* be instantiated. An unpropertied Tommy – a Tommy who does nothing, feelings nothing, has no features – is a surd. What must be instantiated is a happy Tommy, an anxious Tommy or a smoking Tommy.

Before we further discuss the legitimacy of these properties – whether they

exist and whether we can use them if they do in fact exist – let's look at the apparatus that we've been developing. Once again, let S be a set that contains all and only A, B, and C[390]; and remember that we have identified S with 1 (the proposition that Tommy is smoking).

First of all, if C is instantiated, so are B and A. Further, if C, and therefore B, and A, are instantiated, then, and only then, is 1 true. It follows that 1 is true if, and only if, A, B, C are each instantiated. This suggest, though it may fall short of proving, that 1 *is* S that and that 1's being true *is* S's members' being jointly instantiated.

Here we come to some crucial points. It is true that 1 is true if, and only if, all of S's members are instantiated. But there is another reason to identity 1 with S and 1's truth with S's members' being instantiated. 1 and S have comparable decompositions. 1 has the following constituents. Tommy; smoking; Tommy's smoking. S has the property of being identical with Tommy; the property of smoking; the property of being a smoking Tommy. (Actually, each of 1 and S has more constituents than this. We will deal with these remaining constituents in a moment; and when we examine them, we'll see more evidence 1 and S stand in a sufficiently natural correspondence to each other that we are guilty of no artifice in identifying them.)

If 1 is true, that entails that Tommy exists, that something smokes and, of course, that Tommy smokes. If S's members are instantiated, those very same propositions hold.

1 and S have other constituents in common. For example, 1 has as a constituent[391] the function *Tommy does something* (or *Tommy does x*) and also *something does something* (or *x does y*). And S, we may suppose, comprises the property of being something that does something (*x does y*) and also the property of being a Tommy that does something (*Tommy does*

x). So the fact that 1 entails *something does something* is neatly reflected in a fact about S, namely that one of S's members is *the property of something's doing something*. So just as P, when true, entails *something does something*, so S, when its members are instantiated, instantiates the property of *something's doing something.* To echo what we said earlier: what P does in the way of entailing, S does in the way of instantiating. And this, of course, corroborates the view that propositions are set and are true when they members are instantiated.

The problem of synthesizing two propositional constituents.

Everybody knows that, in 1, the two simple constituents, which are (or are similar to) Plato and the property of snoring, are "synthesized" in some way which forms a something truth-evaluable – something that is therefore radically unlike the meaningless heap "Plato, the property of snoring." Frege used the term "saturation" to denote this special process – to denote this I know-not-what that distinguishes propositions, with their characteristic internal unity, from semantically unstructured rubbish, like "Plato, the property of snoring." But, as Davidson observed, the term "saturate" only identifies the problem, and fails to solve it. We are in a position to step into this breach and to say with some precision what this saturation amounts to. Let p_1 be property of being identical with Tommy, and let p_2 be the property of smoking. Now try to find some property p_3 such that any instance of p_3 is an instance of p_1 and also of p_2. So any instance of p_3 smokes and is Tommy.

I'm going to make a bold suggestion. In at least many cases, the "synthesizing" (or "saturating") that occurs in propositions, by means of which (relatively) simple constituents are combined into complex ones, are

instances of exactly the operation just described: given two simple properties p_1 and p_2, a third property p_3 is identified each one of whose instances is an instance both of p_1 and p_2.

Here one might ask: 'Why not just identify

1. Tommy is smoking

with

C The property of being a smoking Tommy?

After all, 1 is true iff C is instantiated.'

Here's why not. Consider an instance of the property of being a smoking Tommy. Obviously that instance will *involve* Tommy smoking. But it will not, at least obviously, contain a discrete part corresponding to Tommy or a discrete part corresponding to smoking. And what is true of an instance of that property is, quite possibly, true of the property itself. It cannot be taken for granted that C has a *discrete* part corresponding to Tommy or to smoking or to anything else. In fact, being a non-spatiotemporal entity, C won't have discrete parts in any straightforward sense at all. And the impression that C is part-less, or lacking in the right kind of parts, is reinforced when we look at what an instance of C is like. Such an instance is, to adapt Kenneth Taylor's apt expression, a dent or ripple in the quantum; it is a shift in the movement of mass-energy. Of course, that mass-energy is (by hypothesis) sufficient for the existence of Tommy and of smoking and, indeed, of a smoking Tommy. But you would have a devil of a time isolating some part of that distribution that corresponding to Tommy, distinguishing that from some other part

corresponding to smoking, and then finding a third that fused those two, otherwise discrete entities into a discrete instance of a smoking Tommy.

By contrast, the proposition

1. *Tommy is smoking*

is, of its very nature, nearly parsed into Tommy, the property of smoking, and a third constituent that synthesizes the two just mentioned. The neat articulations found in 1 are found, in some form, in S. Those articulations, though found in S's membership has a whole, are not found in C taken by itself, making S, but not C, a suitable mirror for the decompositional structure of 1.

States of affairs (or 'facts', as they are sometimes called) do not, as Ian Hacking long ago put it, have the same kinds of "articulations" as propositions. The quantum-ripples on which the truth of a proposition supervenes don't have anything like the structure of that proposition. In fact, the concept of structure isn't applicable in the case of the former. Propositional structure consists of discrete entities that are "synthesized" (in a way that we tried to describe a moment ago) into new entities by means of recursive operations. Actual states of affairs, actual mass-energy coursings, obviously don't have such a structure. (Obviously they have *some* kind of structure. But that structure is identified by giving the locations and forces and of their fundamental particles, not by describing some of their parts as outputs of functions, in the Cartesian sense, of some of their other parts.)

There is a corollary. Consider the proposition:

2. Tommy is snoring.

Let D be the property of being a snoring Tommy. It is obvious that

1. Tommy is smoking

and

2. Tommy is snoring

have much in common. They have in common the constituent *Tommy* (or, in any case, something corresponding thereto). They have in common the functions (or, if you prefer, properties) *the property of being Tommy who is doing something* and *the property of being a something that is doing something.*

In light of this, consider some instance of D. In other words, consider the kind of property-instance, the kind of mass-energy distribution, on which the truth of 2 supervenes. That instance will, of course, suffice for the existence of Tommy, as will an instance of C. But will those instances, in virtue of that fact, have some *discrete* part in common? Unlikely. In neither case do we have Tommy *simpliciter*. We have *Tommy smoking (while sweating profusely and worrying about his algebra test and his nascent nicotine addition, all the while throwing a baseball to his cousin Albert*); in the other we have *Tommy snoring (while having terrible dreams about his future as a second string pitcher on a mediocre time).* There is no such thing, in either case, as Tommy *simpliciter.* An unpropertied Tommy is, to use our old expression, a surd. There is only Tommy doing this or that, having this or that property. A corollary is that two situations that comprise Tommy do not, in virtue of that fact, have in common some discrete, isolable entity. So while it is true that C and D both comprise Tommy[392], the one does not, in comprising that

thing, comprise anything which has anything other than a very rough counterpart in the other.

But now we must consider what we've said about the composition of 1 and, by given obvious extensions of what we've said about 1, about 2. 1 is a set S whose members are:

a: the property of being Tommy
b: the property of smoking
C: the property of being a smoking Tommy

And, by similar reasoning, 2 is a set S* whose members are:

a: the property of being Tommy
b*: the property of snoring
D: the property of being a snoring Tommy

S and S* *do* have well-defined, discrete parts in common. Indeed, they have in common precisely what is had in common by

1. Tommy is smoking

and

2. Tommy is snoring.

If we were to be really precise about it, we'd say each of S and S* comprised, not only a-c, but also

d. the property of being a Tommy that does something (i.e. the property being a Tommy that does something or, in any case, has some property),

and also

e. the property of being a something that does something (i.e. the property being a something that does something or, in any case, has some property),

Let's look at 1 for a second. It immediately and formally implies:

i Tommy smokes;
ii something smokes;
iii Tommy does something;
iv Tommy exists;
v something does something.

And 2 immediately implies:

i* Tommy snores;
ii* something snores;
iii Tommy does something;
iv Tommy exists;
v something does something.

Now let's look at S and at S*. The members of S are:

I The property of being Tommy
II The property of snoring

III The property of being a Tommy that snores.
IV The property of being a Tommy that does something.
V The property of being a something that does something.

The members of S* are:

I The property of being Tommy
II* The property of smoking
III* The property of being a Tommy that smokes.
IV The property of being a Tommy that does something.
V The property of being a something that does something.

Having identified 1. with S and 2 with S*, everything we know about those propositions – about what they entail, about what their components are, and about how their higher-level components are "forged" out of their lower level components – is accounted for.

Another desideratum satisfied by our analysis

Many semanticists today hold that Plato and Socrates are *veritable* objects are constituents of propositions that concern them. So according to Kaplan (1989),who is following and agreeing with Russell (1903), Timmy – the person, not some Fregean concept thereof – is a constituent of both 1 and 2. Other adherents of this view are Scott Soames, Robert Moore, Jeff King, John Perry, Joseph Almog, and (to my knowledge) Michael Thau. (Opponents are J. Searle, M. Dummett, and – with some very heavy reservations to be stated later – the present author.) There are two problems with this view. First, it makes it hard (though by no means impossible) to

deal with Hesperus-Phosphorous problems.[393]

But there is another problem. Consider, not 1 or 2, but rather

3. Plato snores.

At this point in time, Plato doesn't exist. So if he's a (necessary) constituent of 3, then 3 doesn't exist either. Both hard-line direct reference-theorists, Salmon and Soames deal with this by taking a desperate measure, saying that 3 exists *and* that its existence constitutively depends on Plato's existence, *and* that that these two can be reconciled with each other --- by saying that 3 exists, albeit as a non-existent proposition.

This is not necessary or plausible or helpful. 3 does exist. Plato does not. But the property of being identical with Plato, whatever it is, does exist. What is it to be Plato? To be Plato is, perhaps, to be a causal sequence that is initiated by the fertilizing of a certain egg by a certain spermatozoa. To be Plato is, perhaps, to have a soul of a certain kind. What we know is that, if something is Plato, it's being Plato supervenes on something probably having to do with the mass-energy distribution in the cosmos – on its mode of origination, on the quotient of godliness in the particle-interactions that constitute it, on who knows what. In any case, even though Plato does not exist, the property of being identical with Plato does; and that's why we can meaningfully say that Plato does not exist. For, in making such a statement, we aren't picking out a non-object – which would, for that reason, cease to be a non-object – and then saying of it that it doesn't exist. We are saying of an existing property, that of being identical with Plato, that it is uninstantiated.

But our analysis is still consistent with the heart of direct-reference-theory, this being the idea that, for some x such that x is identical with Plato, the proposition meant by "Plato snores" is true just in case x snores, it being

completely irrelevant what properties (other than being a snorer) x has. (The Fregean indirect-referentialist holds that for “Plato snores” to be correct is necessary that, for example, a unique great philosopher to die of hemlock-poisoning snore.) It is clear why our analysis is consistent on this matter with the viewpoint of the direct referentialist. According to our analysis, 3 is true exactly if the following properties are instantiated:

a. the property of being identical with Plato
b. the property of snoring.
c. the property of being a Plato who snores.
d. the property of being something that snores.
e. the property of being something that does something (or has some property).

It is readily seen that, on our analysis, there is some x, such that x is Plato, such that 3 is true exactly if x snores, it being irrelevant what other properties x has. This is exactly what the direct-referentialist holds. At the same time, our analysis guarantees the existence of the proposition *that Plato snores* -- after all, the *property* of being Plato exists, even though it is now uninstantiated. So our analysis of *that Plato snores* gives the direct-reference-theorist the truth-conditions that he wants and that (given certain compelling arguments due to Kripke 1980 and Soames 2003, 2004) he probably deserves. At the same time, our analysis doesn’t put the direct-reference-theorist in the awkward, logic-bending position of having to say that the proposition that Plato snores exists despite the failure of existence of entities on whose existence its own existence is constitutively necessary.

III. *Molecular propositions*

Our analysis applies to molecular propositions, no less than to atomic propositions. But making this transposition go through will involve a certain amount of sensitivity to details and technicalities. A certain amount of "engineering", as opposed to philosophy "proper", will be needed.

Existence-claims: an introduction

In keeping with tradition, I will treat generalizations as molecular propositions. So

4. something snores

is an existential generalization. Given what we've said, the most obvious analysis of 4 would be to say that 4 is the property of snoring (or a set whose sole member is that property) and that 4's being true is identical with that property's being instantiated.

That would not be the correct analysis. Even though 4 is true just in case the property of snoring is instantiated, 4 is not identical with that property and its truth is not identical with that property's being instantiated. We know from Frege that 4, unlike the property of snoring, has a rather distinctive decompositional structure, and that 4 thus consists of (at least) three or four constituents. We also know from Frege[394] what the first two constituents are:

a. the property of snoring
b. the property of being instantiated

Using Conceit # 8, it is not hard to generate some (if not all) of the remaining constituents of 4, these being:

c. the property of being an instantiated function each of whose instances snores.
d. the property of an instantiated function each of whose instances has some property or other.

Let S be a set containing all and only a-d. 4 will be true iff a-d are instantiated. So, from the viewpoint of truth-conditions, there is no barrier to identifying 4 with S or to identifying 4's being true with S's members' being instantiated. Further, 4's structure is easily mapped onto S's. 4, if true, entails that there exists some property that has some instances; that there exists at least one snorer; that at least one property is instantiated; and that at least one property exists. And all of this is precisely what S's members' being instantiated involves.

Now consider the proposition:

5 Something smokes.

By reasoning similar to that just given, the constituents of this proposition are:

a* the property of smoking
b the property of being instantiated
c.* the property of an instantiated function each of whose instances smokes.
d. the property of an instantiated function each of whose instances has some

property or other.

We know pre-theoretically that 4 and 5 have certain constituents in common (e.g. the property of being instantiated or, in any case, *some* property expressed by "something"). Our analysis of 4 and 5 validate everything we know, theoretically and pre-theoretically, about the decompositions and logical forms of the propositions.

Before considering "nothing snores", I want to consider a negative, but non-quantified, proposition:

6. Timmy is not smoking,

in other words,

6. It is not the case <u>that Timmy is smoking</u>.

(The reason for the underlining will be clear in a moment.) Here we need to remember Conceit #7. The two immediate constituents of 6 would *appear* to be "It is not the case" and "that Timmy is smoking." But there's a problem here. We have said that propositions are sets of properties that are true when instantiated. In any case, this is what we have said about atomic propositions and, for uniformity's sake, this is what we must say about molecular propositions. But if we say that, in 6, the underlined proposition is instantiated, i.e. that all its members are instantiated, then *that Timmy is smoking* will be true and 6 will be false and indeed self-negating.

Not-P is true just in case P is false. So "not" can be though as expressing a function that assigns truth to all and only false propositions or (what, in this context, is scarcely different) as a property of all and only false propositions.

For the reason given a moment ago, we don't want to say that *that Timmy is smoking* is a constituent of 6, and will instead say that the property of being *identical with* that proposition is such a property. We will also say that the property of being a false proposition is such a constituent. Then, following Conceit #8, we synthesize those two components into: the property of being the false proposition that Tommy is smoking. That property, it will be noticed, is instantiated exactly if the proposition that Tommy is smoking is false. So, according to our analysis, the constituents of 6 are:

a. the property of being a false proposition
b. the property of being identical with the proposition *that Tommy is smoking*
c. the property of being identical with a proposition that is false and that is identical with *that Tommy is smoking.*

Let S be the set containing all and only a-c. Let us consider what happens when each of S's members is instantiated. First, suppose that c is instantiated. In that case, there is an instance of a proposition that is false and that is identical with the proposition that Tommy is smoking. So 6 is true just in case c is instantiated. Since, if c is instantiated if the other members of S are instantiated, it follows that 6 is instantiated if, and only if, each of the members of S is instantiated. Thus, there are, as of yet, no bars to identifying 6 with S and 6's truth with S's members' being instantiated.

Next, suppose that b is instantiated. In that case, the property *Tommy is smoking* exists -- but it isn't necessarily true. And if 6 should be true – which, according to our analysis, would mean that all of a-c were instantiated – we'd be stuck with the *existence,* but not the *truth of* the proposition that Tommy is smoking. Which is precisely what we want.

Finally, suppose that a is instantiated. In that case, the property of being a

false proposition is instantiated. And, of course, if 6 is true – if, in other words, the members of S are all instantiated – then that property is instantiated. Which, again, is precisely what we want. So our analysis most of our intuitions about the decomposition of 6 (the one exception that being that, where our pretheoretic intuition wants to *that Tommy is smoking* in 6, our analysis puts the property of being identical with that proposition in its place), and is inconsistent with the logical properties of that sentence.

Let us now return to quantified propositions. Consider:

7. Everything snores.

By reasoning similar to that already given, the main constituents of this are:

a the property of snoring
b the property of being universally instantiated
c the property of being identical with a universally instantiated property that is identical with the property of snoring.

Notice that, as always, we just used Conceit #8 to generate the third constituent. c is a property each of whose instances is an instance both of (a) the property of snoring and (b) of the property of being universally instantiated.

Let us now consider:

8. Nothing snores.

I will treat this is as the negation of 5. I will treat it, in other words, as:

8: It is not the case that something snores.

For reasons now already clear, the constituents of 8 include (but are not confined to):

a the property of being a false proposition
b the property of being identical with the proposition that something snores
c the property of being identical with a false proposition that is identical with the proposition that something snores.

Notice that, if c is instantiated, so are b and a. And notice that, if c is instantiated, then the proposition that something snores is false. Finally, notice that c is generated from a and b by conceit #8. Any instance of 6 is an instance both of (a) the property of being a false proposition and (b) the property of being identical with *that Tommy smokes*.[395]

Non-quantified molecular propositions

Let's finish up by considering two non-quantified molecular propositions. First,

9 Tommy is smoking or Tommy is snoring.

Vaguely in the spirit of Montague (who was vaguely in the spirit of Frege), I see 9 as basically saying of a certain pair of propositions (that pair consisting of *that Tommy is smoking* and *that Tommy is snoring*) that at least one its members is true. I thus see 9 as consisting of:

a the property of disjunction, which in this context we may regard as the property of being a pair of sentences at least one of which is true,

and

b. the property of being a pair of propositions, one of them being *that Tommy is smoking*, the other being *that Tommy is snoring.*

Of course, 9 could be false even if both a and b were instantiated. (This would be the case if both *Tommy is smoking* and *Tommy is snoring* were false, but some one member of some other pair propositions were true.) So, using Conceit #8, we look for a third constituent. That constituent must be a property c such that each instance of c is (a) an instance of a pair of sentences, at least one of which is true, and (b) an instance of a pair of propositions, one them being *that Tommy is smoking*, the other being *that Tommy is snoring.* So our final properties are:

a the property of being a pair of sentences, at least one of which is true;
b. the property of being a pair of propositions, one of them being *that Tommy is smoking*, the other being *that Tommy is snoring*;
c. the property of being a pair of sentences, at least one of which is true, and at least one of which is *that Tommy is smoking* and at least one of which is *that Tommy is snoring*.

Of course, if c is instantiated, then so are b and c. And if c is instantiated, then either Tommy is snoring or Tommy is smoking. So c, and therefore a-c, are instantiated exactly if 9 is true. So if S is the set containing a-c, then 9 is

true iff all of S's members are instantiated.

We've thus made it clear (i) what it is for something to be a "constituent" of a proposition", (ii) what it is for an argument to "saturate" a function, (iii) more generally what it is for two propositional constituents to "synthesize" into a third constituent, (iv) what it is to be a proposition, and (v) what it is for a proposition to be true. And our analysis is, with some minor exceptions, consistent with what pretheoretic intuition and common sense have to tell us about propositions.

Section III Alternative analyses

According to Russell (1903), Salmon (1986), Kaplan (1989). R. Moore (1995), J. Perry (1996), J. King (1996, 1997), to name but a few, a proposition is some kind of "structure" consisting of "objects and "properties." So, according to this view, "Socrates is bald" is some kind of structure consisting of Socrates – the man himself, not some property that he has or some concept that he satisfies – and the property of baldness. Little if anything is said beyond that. It is usually said that the structure is not an *unordered* pair – that it is not (Socrates, baldness) – but has some kind of ordinal structure and is therefore <Socrates, baldness> or < the relation of having <Socrates, baldness>>, or some such.

Two questions arise.[396] The ordered pair <2,7> isn't true or false. Neither is the ordered pair <2,7 <less than>>. Given that these ordered pairs aren't true or false, why is <Socrates, the relation of having <baldness>> true or false? There's no denying that we could *use* this as a expressing the proposition now expressed by the sentence "Socrates is bald." We could obviously invent a code whereby "<the relation of having <Socrates,

baldness>>" meant *Socrates is bald.* But this fact – the fact that "Socrates is bald" can be expressed by this new notation -- doesn't tell us anything about the structure of the corresponding proposition that we didn't already know from the fact that it can be expressed by "Socrates is bald."

There is another concern. Obviously the brackets are meant to be symbolic of *some* kind of order that holds among Socrates and baldness (and the relation of having). But what is that ordinal relation? If that ordinal relation bears no resemblance to the relation that would *actually* be had, by Socrates, with respect to the property of baldness, in a situation where he was bald, then it wouldn't be clear, at least not obviously, why what relation was the right one: some kind of reason – the likes of which have never been provided by advocates of this model – would have to be provided why this particular ordinal relation (whose nature, by the way, has not, to my knowledge, been disclosed) was the right one. So, thus construed, the analysis in question is highly programmatic. On the other hand, if that relation *is* the very relation that would be had by Socrates, with respect to the property of baldness, in a situation where he was bald, then Socrates' baldness – his actual baldness – would be built into the very existence of the proposition that he was bald. In other words, the proposition that Socrates was bald would, by virtue of its own configuration, make him bald, and would thus predetermine its own truth. But obviously the truth of *Socrates is bald* isn't self-determining. Basically, if propositions model facts *too* well, then they force contingent propositions to be true. If we turn the proposition *Socrates is (or was) bald* into a model consisting of Socrates standing in the relation of having with respect to baldness, then the proposition *itself* is the fact of Socrates' being bald, and no way is left open for the obvious possibility that that proposition might well be false.

There is another problem with the view in question. Though is now false,

the proposition *Socrates is alive* obviously exists. In any case, it would be very artificial to maintain that that proposition didn't exist. What if, unbeknownst to you, your friend Joey died some time during the last five minutes? Surely you wouldn't want to say that the proposition *Joey loves to read* suddenly ceased to exist. Apart from an acceptance of a certain version of direct-reference theory, there is no reason to accept that view.

According to the view that I am advocating, what is a constituent of the proposition *Joey likes to read* is the property of being identical with Joey. [397] And that property doesn't go away even when Joey passes on (it only ceases to be instantiated), leaving my analysis without the problem had by the Russellian analysis just mentioned.[398]

Frege held it wasn't Socrates *per se* that was a part of the proposition *Socrates is alive* but was instead some concept that Socrates uniquely satisfied. (Frege's analysis is therefore consistent with the fact that "Joey likes to read" holds onto its meaning even if Joey ceases to exist.)[399] But, apart from that, Frege's view is quite as vague as Russell's. Frege was extremely opaque about the nature of the structure of this proposition; and what he failed to provide in the way of clear formulations he tried, unsuccessfully, make up for with rather unhelpful terms such as "saturate." .

What is good about the Russell-Frege-Kaplan-King view is that it insists that propositions resemble the sentences that express them. So that view is, at least, consistent with the fact that languages can be learned and used (after all, learning and using languages involves there being systematic ways of hooking up meanings to sentences). And, as we'll now see, other analyses don't do such a good job of satisfying this requirement.

For a while, it was widely held (by e.g. Clarence Lewis and Rudolph Carnap) that the "intension" (what we would now call a "meaning" or

"proposition") of a sentence was identical with the class of all propositions that could be inferred from that sentence. So (a natural extension of this view would be that) the proposition *John knows that corn is yummy* is the class of all propositions deducible from that proposition, e.g. *John believes that corn is yummy, John knows something, John has some kind of attitude towards the proposition that corn is yummy.* There are a few problems with this view, some minor, some major. A minor problem is that, as it was stated by Lewis and Carnap, it is viciously circular, since many propositions that can be deduced from *John believes that corn is yummy* that involve that very proposition, e.g. *John believes that corn is yummy* and *either John believes that corn is yummy or there are square circles.* This particular problem is not, I believe, insuperable, since the position can be amended to avoid it. The right emendation would seem to be: A proposition is identical with the class of propositions that don't comprise that proposition that can inferred from it.

Even so, the position in question isn't very plausible. Consider the class of all propositions that follow from *John believes that corn is yummy*. That class doesn't have a structure corresponding even remotely to the structure that we would expect that proposition to have, give that it almost certainly has a structure similar (in some non-trivial ways) to the sentences that express it. Also, in order to figure out whether a given proposition is a consequence of *John believes that corn is yummy*, one must, it seems, grasp that proposition itself. The set of propositions that follow from *John believes that corn is yummy* is infinitely large. One can grasp that set only in the sense that one grasps some rule that generates its members – obviously one cannot grasp them all one by one. And grasping that rule would involve grasping *John believes that corn is yummy* or some other similar proposition.

Now we turn to what is, these days, perhaps the most popular analysis of propositions and, indeed, of modality: the possible worlds analysis.

According to possible worlds semantics (PWS) a proposition is necessary if true in all (possible) worlds, and a proposition is a function that assigns truths-values to worlds or, alternatively, is a set of worlds. So the proposition *grass is green* is a function that assigns *truth* to all the world where grass is green, *falsity* to worlds where grass is not green, and either assigns falsity or no truth-value to grass-free worlds.

In this context, the first question to ask is: what is meant by the term "world?" Oftentimes, advocates of PWS say that "worlds", as they are using the term are propositions (or sets of propositions). In that case, as David Lewis (1986) was quick to point out, the PWS slogan that "a proposition is a set of worlds" becomes "a proposition is a set of propositions", which is not only false but viciously regressive, the same thing *mutatis mutandis* holding of the statement "a proposition is a function that assigns truth-values to worlds." (This is one of the reasons Lewis, an advocate of PWS, chose to use the word "world" literally.)

Some advocates of PWS – aware that treating worlds as propositions (or sets thereof) eviscerates PWS – say that worlds are just that – worlds: concrete objects like our own, "maximal states of affairs." *Not – nota bene* – properties that, *if* instantiated *would* be worlds. But actual worlds.

Here we run into several problems. First, not too many are going to grant the existence of *bona fide* alternative worlds. Second, even if they do exist – and maybe they do (who knows?) – we can't know anything about them *except what we can figure out on the basis of modal knowledge obtained independently of our knowledge of those worlds.* In *On the Plurality of Worlds*, Lewis says, rightly, that we don't know about alternative worlds by taking a tour of them. We know about them by considering materials that are available to us here and "employing a principle of recombination" – by, basically, applying modal principles that we already

know to hold and then seeing if those principles square with the proposition whose modal status we wish to know (e.g. there might be square circles in other worlds). But in that case, the worlds answer to us, not us to them. Our modal knowledge is the tail that wags the dog of these other worlds; and from the viewpoint of explaining (or validating) our modal knowledge, those extra worlds are useless (though there may be other good reasons for positing them).[400]

There are other problems with PWS (considered as an analysis of what propositions are). The class of worlds where *1+1=2* is identical with the class of worlds *triangles have three sides*. But those propositions are obviously distinct.[401] That's you can think the one without thinking the other. (And there's no intensional fallacy here, since propositions *are* intensions.) In its most primitive form, PWS can't distinguish analytically equivalent propositions. Some advocates of PWS deal with this by saying the following:

There are two ways of thinking of what functions are. A function can be though of as a a set of ordered pairs. So, from that viewpoint, the function F(x)=x+1 just is the set of pairs: <1,2>, <2,3>….And from that viewpoint, F(x)=x+1 is identical with the function F(x)=x+(90-89). This is the *extensional* view of functions: functions just *are* the pairs they generate – *how* they generate them is irrelevant. But extensionalist view of functions isn't the only one, since there's also the *intensionalist* view. The intensionalist says that functions are individuated by *how* they generate the pairs they generate. So, for the intensionalist, for functions to be identical, it is necessary but not sufficient that they generate the same ordered pairs. To avoid falsely identifying distinct, analytically equivalent propositions, PWS must take an intensionalist view of functions. So propositions are functions,

and those functions are individuated by their intensions, not their extensions.

There are two problems here. First, this view still demands that there be other worlds. If those worlds are propositions, then we're no further along. If those worlds are concrete, then we're stuck with some dubious (and, I believe, demonstrably useless) metaphysical lumber. But leaving those problems aside, the solution just proposed collapses on itself. Propositions *are* intensions (though not all intensions are propositions, e.g. the propositional function *x is a human being* is not a proposition, even though it is an intension). This by itself is a problem for PWS since much of its *raison d'etre* is that it replaces intensions, which for various reasons are seen as inimical to scientific thought, with extensions, which are seen as methodologically unimpeachable. So in taking an intensionalist view of functions, PWS is undermining itself. For, in taking such a view, PWS is letting properties back in – properties are no longer to be identified with sets of their instances. But, by parity of reasoning, propositions needn't be identified with sets of *their* instances, i.e. with sets of worlds where they hold. So PWS cannot, without arbitrariness, take an intensionalist view of functions but it cannot, without failing to distinguish between obviously distinct propositions, take an extensionalist view of them either.

Here's a similar problem. Consider the proposition *grass is green*. As we know, PWS identifies that proposition with a function (that assigns truth to worlds where grass is green) or with a set of worlds (containing all and only those worlds where grass is green). If it isn't thought of as a set of ordered pairs, that function must be thought of as a *rule* of some kind. What kind of rule? One that assigns truth to certain worlds on the basis of their satisfying some condition. What condition is that? The condition that, in those, grass is green. So the function in question, understood intensionally,

must be understood *in terms of the very proposition in question*. Thus, unless the advocate of PWS wants to falsely identify *1+1=2* with *triangles have three sides*, he must take an intensionalist view of functions, which means that he must circularly understand the functions with which he aspires to identify propositions with the very propositions that aspires to identify them with.

Frege made it clear that propositions are not ideas; they are the *contents* of people's ideas, but are not themselves psychological entities. Since Frege, many attempts have been made to say what propositions are. (Some, e.g. Quine, have, at least at times, denied their existence.[402] But that is a topic for another paper.) Those attempts have either been programmatic to the point of near-emptiness or demonstrably false or both.

But given a certain healthy dose of Platonism, it is not impossible to identify structures that model the properties that know, on both theoretical and commonsensical grounds, propositions to have. Given any proposition P, it is clear that there is some set of properties $p_1 \ldots p_n$ that will be instantiated just in case P is true. Supposing that S is the set containing those properties, this suggests that S *is* P and that P's being true *is* S's members' being instantiated.

We found that, taking this approach, we could give non-metaphorical meanings to terms that, though liberally used, have thus far been used in a purely metaphorical way – terms like "constituent" and "synthesize." A "constituent" of a proposition, we found reason to believe, is a *member* of the set of a properties whose joint-instantiatedness is necessary and sufficient for that proposition's truth. As for the term "synthesize": since all components of propositions are properties, only properties can, in any relevant sense. be synthesized. And two properties p_1 and p_2 are "synthesized" into another component when a third property p_3 is identified each of whose instances is

an instance of both p_1 and p_2. Though no doubt imperfect and incomplete, it seems to me that this conception of what propositions are has sufficient promise to warrant its being considered and, perhaps, its being further developed.

Fodor on Concepts: How to be a non-Atomist without being a Holist

By a "concept", I mean one's ability to think about a thing. I am able to have thoughts about Socrates; therefore I have a concept of the man. I am able to have thoughts about the number four; therefore I have a concept of that entity.

Jerry Fodor argues that, with a few trivial exceptions, our concepts are "atoms". A grasp of any one thing presupposes a grasp of nothing else. I can have a concept of knowledge without having a concept of truth or belief or justification. I can have a concept of causality without have a concept of space or time or persistence. I can have a concept of Socrates without having a concept of corporeality, location in space or time, movement in space, or causality, personal identity, or sentience.

To sum up: For Fodor, one can have a concept of (practically) any one thing without having a concept of anything else. A grasp of (practically) anything presupposes a grasp of nothing. Our concepts are "atoms"; they are indivisible and ultimate; they cannot (typically) be broken down into other concepts.

Fodor provides three ingenious arguments for this astonishing position. The purpose of this paper is examine those arguments. I believe that all three arguments involve non-sequiturs. Where the first two are concerned, the non-sequiturs are, I believe, fairly obvious. But the non-sequitur involved in the third is by no means obvious; and to expose it, we will have to put forth some substantive *positive* theses about the nature of concepts.

§ Let us start with a terminological point. The term "concept" is ambiguous. The word "concept" can refer to some platonic, non-mental

entity, as in: “anything falling under the concept *square* also falls under the concept *closed figure*.” But the word “concept” also has a psychological sense; it can refer to one’s *grasp* of something, as in: “Locke identified concepts with mental images” or “before 1930, nobody had a concept of a quark”.

We need two different words here, if we are to avoid intolerable confusion.[403] To denote concepts in the *non*-mental sense, let us use the word “concept$_O$” (plural: “concepts$_O$”). To denote concepts in the psychological sense, let us simply use the word “concept” (no subscript). [404]

I will talk about “having” concepts and “grasping” concepts$_O$. I *have* a concept (no subscript) of the number two. And I *grasp* the corresponding concept$_O$.

§ Here is a view about concepts that certainly has at least some initial plausibility. It is the view that I hold and that Fodor *rejects.*

Possession of any one concept presupposes possession of others. To grasp any one thing, one must be able to grasp others. One cannot think about, or mentally represent, a given thing unless one think about others. Consider the concept$_O$ of a square. Suppose some creature had no idea what it was for a figure to have four sides, or what it was for an edge to be straight. Would that thing have any idea what it was for something to be a square? Intuitively, the answer seems to be “no”. So to grasp the concept$_O$ *square*, one must grasp the concepts$_O$ of the number four, of straightness, of flatness, and so on.

Of course, what we just said about the concept$_O$ *square* applies to a grasp any one of *these* concepts. Consider the concept$_O$ of the number four. To have a concept of that thing, one presumably has to have a concept of the

numbers three, two and so on. To have concepts of *those* things, one must presumably have concepts of magnitude, order, and who knows what else.

The concept$_O$ of a square might seem like a tendentious one. A square *is* a structure; to be a square just *is* to have certain structural or ordinal features. So given only that a grasp of *that* concept$_O$ presupposes a grasp of others, perhaps no general truths follow regarding the nature of conception.

But what is obviously true of the concept$_O$ *square* is, at least arguably, true of all other concepts$_O$ – even those that *seem* that most devoid of structure. Consider the concept$_O$ of redness. One might naively think that an otherwise conceptless creature could grasp *this* concept$_O$. All the creature needs is the ability to have a certain kind of phenomenal content – it doesn't need an articulated network of background concepts.

But such a view is dubious. An object that really *is* red -- i.e. that has the microphysical properties common to stop-signs, apples, and the like – can be given to a person of good eyesight through very different sorts of phenomenal contents. You have 20/20 vision. You look at the apple from place p1. The phenomenology of your perception of its color is one way. You look at the apple from place p2. Now the phenomenology of your perception of its color is different. This is the case even if you are looking at the apple under optimal conditions – you're in a brightly lit room, your eyes in perfect working order, and so on. No matter how "optimal" or "standard" the conditions of perception are, there will never be a one-one correspondence between things' actual colors and your phenomenology. A given shade of red will given to you through different phenomenal contents.

Any artist knows this. If you want to depict its color as it appears to you from p1, you use one mix of paint. If you want to depict its color as it appears to you from p2, you use a different mix of paint. This is the case, as any artist knows, not just if you are in a room with bad or uneven lighting: it holds even

in circumstances where the lighting conditions are optimal – where they are unnaturally good.

To see something *as* having a certain color isn't just to have an experience with a certain phenomenology. When you just *see*, so to speak, that something is red, a wealth of background information is operative. Phenomenology is certainly part of the story, but not all of it. The phenomenology has to be appropriately contextualized. And this contextualization seems to presuppose some understanding of the conditions of perception; it thus presupposes an articulated picture of the world. So even a concept of redness – surely our best candidate for a "simple" concept, a concept that doesn't presuppose a wealth of other concepts – is available to only to a creature with a formidable representational endowment.

What about *purely* phenomenal concepts? Consider the experience of tasting honey. Consider what it is like to have *that* kind of experience. Perhaps an otherwise conceptless creature could have a concept of what it is like to have *that* kind of experience. So here, it will be thought, we have found a genuinely simple and foundational concept.

But even here there is room for doubt. First of all, there is a difference between, on the one hand, *being able* to have an experience with that sort of phenomenology, and on the other hand, having a *concept* of that sort of experience. Perhaps the ability to *have* that kind of phenomenal content doesn't require any sort of conceptual background. But that doesn't mean that the same is true of the ability to have a *concept* of that kind of experience. Having a *concept* of that kind of experience would seem to involve being able to *recognize* that occurrence of states with that sort of phenomenology and also have thoughts about such states. Even though (perhaps[405]) little or no cognitive wherewithal is needed to *have* the phenomenology, it is hard to believe that the same is true of the ability to *recognize* the phenomenology

for what it is. To recognize is to single out. Singling a thing out seems to involve knowing how that thing is different from other things. So, at least on the face of it, recognizing a thing involves having a concept of how thing differs from other things. So having a *concept* of a certain kind of phenomenology would seem to involve being able to compare it with other things, and thus to presuppose that one has concepts of those other things. So while the ability to *experience* a certain kind of phenomenal content may be conceptually innocent, the same appears not to be true of the ability to have *concept* of that sort of experience.

Let us state what this extremely brief survey has told us. There does not appear to be *any* concept that we can have without having others. There does not appear to be any concept_O that one can grasp without grasping many others.

But there is more to say. Obviously we don't just grasp concepts_O; we also grasp spatiotemporal objects – rocks, trees, and the like. Let us briefly talk about our concepts of these things.

We know about external objects entirely through sense-perception. A sense-perception necessarily presents things *as* having various properties – as having a certain color, shape, location, state of motion. Strictly speaking, what is given to one in sense-perception is never (just) *a thing*: it is a state of affairs. You don't just see *the rock.* You see a state of affairs involving the rock: you see *an* object of a certain shape, color, relative location, and so on. If a thing is to be represented to you as having a certain color, location, state of motion, and so on, it seems to follow that – at some level, in some way – you have cognitive access to these properties: concepts of these properties constitute your representational palette.

So any sense-perception has a great deal of "descriptive structure", and also presupposes some kind of cognitive access to a wealth of properties.

Since sense-perception is essentially involved in our being able to think about external objects, it would seem to follow that our concepts of such objects are replete with conceptual structure.

To sum up, there is at least something to be said for the following view. A concept of a concepts$_O$ always seems to presuppose networks of background concepts. The same is true of our concepts of external objects. Attempts to find genuinely foundational concepts – concepts an otherwise conceptless creature can have – seem to bear no fruit. Our best candidates for that honor (e.g. the concept$_O$ of what it is for something to be red, the concept$_O$ of what it is to have an experience with such and such phenomenology) turn out to bristle with exactly the sort of structure we were hoping they wouldn't have.

So, on the face of it, there seem to be no *atomic* concepts – no concepts that are not to some degree constituted by other concepts. There seems to be no concept C such that a creature could have C without also having various other concepts. There are no foundational concepts. After all, a foundational concept is one such that possession of it doesn't presuppose possession of others.

§ In *Concepts*, Jerry Fodor argues for the exact opposite of this view. He argues that, with some trivial exceptions, all concepts are atoms. For any concept C, possession of C does *not* presuppose possession of other concepts. In principle, one could have a concept of knowledge without a concept of belief or truth (this is one of Fodor's own examples); of justice without having a concept of sentience; of existence in the external world without having a concept of persistence in space and time or of causality; of Tuesday without having a concept of Monday (again, one of Fodor's own examples).

Fodor's overall view is generally rejected. Most hold that atomism is false. But it has not yet been said, to my satisfaction, just where his

arguments for that view go wrong. In this respect, Fodor is a latter day Berkeley. Everybody thinks that Berkeley is wrong: *of course* there are mind-independent objects. But it is not often that one hears it clearly stated just where Berkeley's reasoning goes astray. Similarly, no one has, in my view, pinpointed the non-sequiturs in Fodor's arguments, though many take it for granted that his conclusion is false.

In this paper, I will present Fodor's three arguments, and I will try to show where they go astray. One those arguments involves a fairly blatant non-sequitur. But the other two are quite sturdy. Identifying the fallacies in them involves producing some positive alternative theories about the nature of conception, and is therefore not a zero-sum affair.

§ First of all, Fodor actually uses the term "conceptual atomism" to refer to two distinct (though related) doctrines.

(i) One's *grasp* of a given thing does not typically presuppose a grasp of other things. No concept presupposes possession of any other concept.

(ii) Concepts$_O$ (note the subscript) are atoms. They cannot be broken down into other concepts$_O$. Consequently, there are (with some trivial exceptions) no analytic truths.[406]

As we discussed, it certainly *seems* that, if one grasps the concept$_O$ *knowledge*, one also grasps the concept$_O$ *true belief*. This is exactly what Fodor denies. In his view, one *could* grasp the concept$_O$ *knowledge* without

grasping any other concepts$_O$.

Fodor admits that there are *some* exceptions to this. A grasp of the concept$_O$ *red cow* is not an atom, as it plainly consists of a grasp of the concept$_O$ *red* and also the concept$_O$ *cow*. The same is true of the concept$_O$ *prime number*. If the concept$_O$ one is grasping is *patently* synthesized out of other concepts$_O$, then a grasp of it is not an atom.[407] But with these exceptions, one's grasp of a concept$_O$ is always an atom.

Fodor's denial of (i) is connected with his denial of (ii). Fodor thinks that concepts$_O$ (with the trivial exceptions just given) have no internal structure. The concept$_O$ *knowledge* has no constitutive connections to the concepts$_O$ *belief* or *truth* or *justification.* Of course, if Fodor is right about that, then we can see why a grasp of *knowledge* wouldn't presuppose a grasp of these other concepts$_O$.

Fodor obviously admits that concepts$_O$ typically *appear* to have a great deal of analytic structure. Obviously *knowledge is always true* appears to be analytic; and so do *2 comes before 3* and *causes precede their effects.* He goes to great and ingenious lengths to explain away this appearance of analyticity.[408] We will consider these efforts soon enough.

Fodor grants that there are *some* analytic truths. The concept$_O$ *red cow* is "patently phrasal", as Fodor puts it. It is *obviously* built out of other concepts$_O$ (*red* and *cow*). Any such concept$_O$ will clearly be associated with various analytic truths. (*x is a red cow* implies, and is implied by, the conjunction *x is red and x is a cow*.) Fodor grants the existence of analyticities of that kind. And Fodor grants that *x is a cow* entails *x is a cow.* Notice that those analyticities are *trivial.* They don't expose any hidden depths on the part of the relevant concepts$_O$. In Fodor's view, all analytic truths are in the same category as:

(*) *if x is a red cow, then x is a cow.*

We can be a little more precise about this. (Here we are going beyond what Fodor himself says.) In (*), the non-logical terms occur inessentially. Suppose you replace "cow" with "horse". The result will be a truth. Suppose you replace "red" with "green" or "square": what results will be a truth. The truth of (*) doesn't turn on anything peculiar to the concepts$_O$ *red* and *cow.* Any other concepts could take their place. Since (*) doesn't turn anything peculiar to those concepts$_O$, it doesn't really *tell* you anything about them. It doesn't tell you how *those* concepts$_O$ are different from the concepts$_O$ *horse* and *green.*

Now consider the proposition:

(**) *Any case of knowledge is a case of true-belief (but not vice versa).*

If you replace "knowledge" with "justice", or if you replace "true" with "green", what results is a falsehood (or possibly nonsense). So in (*), "knowledge" occurs *essentially*. The truth of (**) turns on something peculiar to the concepts$_O$ *knowledge* and *belief* and so forth.

Fodor grants analytic status to propositions like (*) – propositions whose truth doesn't turn anything peculiar to the concepts$_O$ occurring therein. Fodor *denies* analytic status to propositions like (**) – propositions whose truth *does* turn on the peculiarities of the concepts$_O$ involved. He appears to grant analytic status to propositions that are *formally* true. But apart from that, he seems to think that there are no, or very few, analytic truths. (Fodor seems to think that *bachelors are unmarried males* is analytically true. But even here he has reservations.)

§ Fodor provides three arguments for atomism. Let us turn to Fodor's first of these. Two thousand years of work have yielded no unexceptionable definitions of 'knowledge', 'justice', or 'number'. In fact, even the most pedestrian concepts$_O$ are indefinable. Consider the concept$_O$ associated with the verb 'to paint.'[409] Fodor goes through a number of reasonable-seeming definitions of this concept$_O$, and shows that they all fail. For example, it might seem that 'to paint' something is just to cover it with paint. But this is not so. If you dip your paint-brush in a paint-bucket, you are not painting your brush, even though you are covering it with paint.

Analogous remarks apply to attempted definitions of practically *any* concept$_O$. Try giving a definition of 'knowledge' or 'justice'. You'll probably find that there is a counter-example to it.

Given the apparent indefinability of most concepts$_O$, Fodor infers that they have *no* structure: they are simples. The idea is that, if a concept$_O$ is not simple, then it is built out of other things, presumably other concepts$_O$. If it is built out of other things, then there is obviously some true proposition saying *how* it is built out of those other things: there is a definition or analysis of that concept$_O$. Since there are no adequate analyses of concepts$_O$, it follows that they are not built out of other concepts$_O$.

On the face of it, this argument concerns *concepts$_O$*, not concepts. It shows that concepts$_O$ – not conceptso – are atoms. But it indirectly supports the idea that concepts$_O$ are atoms. If two concepts$_O$ are not analytically related, then it is hard to see why a grasp of the one would involve a grasp of the other. So if *no* two concepts$_O$ are thus related, then it is hard to see why a grasp of any one concept$_O$ would presuppose a grasp of others. The other side of the coin is that if *any* concept$_O$ is individuated by its analytic relations

to various other concepts$_O$, then presumably one must have some grasp of these analytic relations to grasp any given concept$_O$: so one would have to grasp a multiplicity of concepts$_O$ to grasp a single one of them. So given that Fodor wants to defend the view that concepts (no subscript) are atoms, he is probably right to attack the view that concepts$_O$ (note the subscript) are atoms.

Evaluating Fodor's 1st Argument

This argument involves a non-sequitur. Given his premises, what Fodor should have concluded is that concepts$_O$ have an *infinite* amount of internal structure. Fodor has made a case that it is hard to give conceptually$_O$ necessary and *sufficient* conditions for a thing's falling under a concept$_O$. [410] But for practically any concept$_O$, it is extremely easy to give conceptually$_O$ *necessary* conditions. (In effect, Fodor himself shows this.) Conceptually$_O$ necessary conditions obviously correspond to analyticities. For x to be an act of painting, it is conceptually$_O$ *necessary* that x be an act of putting paint on something. This by itself shows that there is an internal or constitutive relation between 'Bob painted y' and 'Bob put paint on y' and, therefore, that the concept$_O$ 'to paint' has conceptual structure – is not an atom. [411]

Fodor goes through several attempted definitions of the concept$_O$ *paint* (the concept$_O$ corresponding to the verb, not the noun, 'paint'). He points out that, for each one of them, there is something that is *not* an act of painting that falls under that definition. What Fodor does not say is that each of these

attempted definitions gives a *necessary* condition for x's being an act of painting. So the evidence suggests that it is hard to give conceptually$_O$ necessary *and sufficient* conditions for x's falling under the concept$_O$ *paint.* But the evidence also suggests that we can generate *ad infinitum* conceptually$_O$ *necessary* conditions for something's being an act of painting. So given that necessary conditions indicate conceptual structure, it follows that *paint* has an *infinite* amount of conceptual$_O$ structure.

This point applies *mutatis mutandis* to practically any other concept$_O$ one cites. Consider the concept$_O$ *knowledge.* It is easy to generate conceptually *necessary* conditions for x's falling under that concept$_O$. Gettierologists have generated countless such conditions. This is not to mention the ones that have always been known: x must be true and justified and a belief. So it seems possible to generate conceptually$_O$ *necessary* conditions *ad infinitum* for x's being a case of knowledge. If that is right, then the concept$_O$ *knowledge* has an infinite amount of analytic structure, contrary to what Fodor says. If that is wrong, then Fodor is again wrong. Suppose that, after adding some more epicycles, Gettierologists produce a correct analysis of knowledge (x is knowledge iff x is reliably caused, justified, true, well-entrenched….belief). In that case, once again, Fodor is wrong to say that the concept$_O$ *knowledge* has no analytic structure.

§ Let us now turn to Fodor's second argument. Right now, let us focus on one's *grasp* of the concept$_O$ *knowledge.* Let us focus one's concept (no subscript) of that concept$_O$.

Suppose that one's concept of *knowledge* really did implicate concepts of *truth*, *justification*, and *belief* – along with whatever concepts are presupposed by *these* concepts. In that case, we would expect cognitive operations concerning *knowledge* to be slower, requiring of more acumen,

and generally more unwieldy than cognitive operations concerning *justification* or *truth.* But we find no such thing. Thought processes involving one's concept of knowledge are *at least* as speedy and energy-efficient as thought processes involving the concepts that – if the non-atomist view is correct – are constitutive of one's knowledge-concept. Thoughts about knowledge are *at least* as speedy and efficient as thoughts about truth, justification, belief, sentience, and so on.[412]

Of course, we must ask: *why* must it be the case that, if a grasp of *knowledge* implicates a grasp of *truth*, *reliability*, *belief*, and so on, then cognitive operations with *knowledge* would be slower than those with (say) *truth*? Here is Fodor's answer. First of all, when it is said that one's ability to think about knowledge 'implicates' or 'presupposes' an ability to think about (e.g.) truth, what is meant is *not* that there is some merely *causal* connection between the two abilities, but that there is some kind of *metaphysical* relation: *what it is* to be able to think about x as being a case of knowledge *is* (inter alia) to see that x has the properties of being a belief that is non-accidentally true and reliably caused …So for there to be some kind of constitutive relation between one's ability to think about knowledge and one's ability to think about justification or truth is for there to be some *metaphysically* necessary relation between these two capacities.

Given what Hume said, it seems that there can be no metaphysically necessary relation between x and y unless x *makes up* y or vice versa. (Identity can be seen as a limiting case of this relation.) If x and y are spatio-temporal entities – and surely concepts are such things, for they are realized by configurations of neural matter – then no necessary relation can hold between them unless one of them literally makes up the other, unless, in effect, one of them is a part of the other. But if one's 998-concept, and one's 997-concept… make up one's 999-concept, then presumably in exercising

the latter, one is exercising the former. Of course, to 'exercise' one's 999-concept just is to think about 999. So it seems to follow that, if there are constitutive relations among one's concepts of 999, 998, 997…then in thinking about 999, one is ipso facto thinking about 998, 997, 996…2 and 1. But, as Fodor says, it seems unlikely that I must simultaneously think about 1000 numbers to think about 999. Therefore, my 999-concept is *not* constitutively related to my 998-concept , my 997-concept …or to any other concept$_O$ of mine to which it might initially appear to be constitutively related.

Evaluating Fodor's 2nd Argument

This is a very clever argument. But it covertly assumes the very atomism that it is meant to establish.[413] What the non-atomist maintains is *specifically* that there are no atomic concepts: there are *no* concepts that are not constituted by other concepts. So there are no concepts "at the bottom" (for there is no bottom). What the non-atomist maintains is specifically that we mustn't think of the body of our concepts as having a structure like that of a building – where there is a ground floor, a second floor resting on the ground floor, a third floor resting on the first two, and so on. What he maintains is specifically that our concepts support one another. The right metaphor is a web or arc, not a skyscraper.

Fodor's argument presupposes that, *if* non-atomism is right, the relation between one's grasp of the concept$_O$ *knowledge* and one's grasp of (say) the concepts$_O$ *truth* or justification is like the relation of the third story of a building to the second story: the one supports the other but not *vice versa.* Fodor seems to think that, if non-atomism is correct, then the relation between

one's grasp of the concept$_O$ and one's grasp of the other is thus asymmetrical. But this is exactly what the non-atomist denies. The non-atomist maintains *specifically* that there are no concepts at the bottom supporting everything else and that everything supports everything. He maintains that, while one's grasp of the concept$_O$ *knowledge* is constituted (in part) by a grasp of the concept$_O$ *justification*, the constitution relation goes in the other direction as well. One's grasp of the one concept$_O$ supports one's grasp of the other – as different parts of a web support one another. There is nothing at the bottom that supports a grasp of everything else. So, if the non-atomist is right, there is no expectation that operations with the one concept$_O$ should be *more* unwieldy than operations with the other. Since one's various concepts (no subscript) hold one another in place, there is a certain equality among them. None of them is heavier (so to speak) than the other.

Fodor's idea is that, if non-atomism were right, then certain concepts would be feather-light, while others would be heavy. There would be those concepts that are fundamental, and thus don't presuppose a baggage of various other concepts: these are the light ones. And there would be those that *do* presuppose this sort of baggage: these are the heavy ones.

But it is just this sort of picture that the non-atomist rejects. What the non-atomist rejects is precisely the idea that there are *foundational* concepts, and then a second set of concepts built on *those*, and a third set built on the second set, and so on. So what the non-atomist rejects is precisely the view that this kind of asymmetry prevails. For the non-atomist, constitution is typically a mutual relation. Given this, Fodor's argument collapses. One's concept of *knowledge* is constitutively related to one's concept of *justification.* But this doesn't mean that *one* of those concepts must be more basic than the other or, therefore, that use of the one demands more sophistication or energy than the other. Those concepts hold *each other* in place; neither is sitting on

top of the other; neither is, in any absolute sense, simpler or more fundamental than the other. So neither is more wieldy or unwieldy than the other. On the non-atomist picture, constitution is (or can be) a mutual relation. Given only that the one concept is constitutively related to the other, no inferences can be made as to whether operations with the one concept are in any sense more demanding than operations with the other.

Of course, when we are talking about *physical* objects, like statues, the constitution relation is asymmetrical. If the x constitutes y, then y doesn't (typically) constitute x. (In the degenerate case where x and y are *identical*, then perhaps the constitution relation can be thought to go in both directions – but only because, in that case, it doesn't really go in any direction.) So *if* the sense in which one's concept of truth were constitutive of one's concept of knowledge were identical with the sense in which the clay is constitutive of the statue, *then* the constitution relation in the first case would probably be asymmetrical. But when we say that certain concepts are "constitutive" of others, we are probably not talking about the kind of constitutive relation that holds between the clay and the statue. The latter is, at most, a metaphor for the former – and it is (I think we are already seeing) a very misleading metaphor.

§ Fodor's 3rd argument is the most compelling.

For the sake of argument, suppose that some kind *non*-atomism is the case. In that case, a grasp of the concept$_O$ *square* presupposes a grasp of various other concepts$_O$ – *four, straight, planar*, and so on. The idea is that to think of something as a square, you must think of it also as flat, four-sided, and so on. So, for the non-atomist, to think of x as a square just *is* for you to accept various other things about x. Put another way, you believe that x is a

square exactly if you accept that x is flat, four-sided, and so on. The content of a belief supervenes entirely on what you accept in virtue of having that belief – on what you are prepared to infer. Thus, for the non-atomist, beliefs are individuated by their "inferential roles". Let C be some brain-state of yours. For the non-atomist, C qualifies as a belief that x is a square exactly if, in virtue of having C, you are prepared to hold that x has four-sides, that x is not a hexagon, that x is planar, and so forth. So C's content supervenes entirely on C's inferential role. Thus non-atomism presupposes what might be called "inferential role semantics" (IRS).

(For reasons I will give, I myself do *not* think that non-atomism requires that some kind of inferential role semantics be correct. Right now I am presenting Fodor's argument.)

Fodor says that IRS unacceptable for a couple of related reasons. Let b1 be my belief that x is a square and let b2 be Gauss's belief that x is a square. Gauss knew a lot things about squares that I do not. So what he was prepared to accept from the proposition *x is a square* is very different from what I would be prepared to accept from that proposition. So b1 has a very different inferential role from b2. For IRS, this means that, contrary to our hypothesis, b1 and b2 don't have quite the same content – at most, they have *similar* content. But obviously they do have the same content: they are both beliefs to the effect that x is a square.

Let us generalize this point. For an advocate of IRS, differences in inferential role mean differences in content: unless there is identity of inferential role, there is not identity of content. Given this, let B and B* be any two beliefs. Unless B and B* have precisely the same inferential role, they do not have the same content. The trouble is that practically *no* two beliefs have precisely the same inferential role. Let Smith and Brown be two mathematicians who are in exact agreement as to the mathematically

significant properties of the number two. Given *n=2*, Smith will, whereas Brown will not, accept the proposition *n=the age of my son when he learned how to read*; and Brown, whereas Smith will not, accept the proposition *n=the number of pages of my most impressive proof*. So, for IRS, the two beliefs in question *don't* have the same content. But obviously this is false – obviously they have precisely the same content: they are both beliefs to the effect that n=2. For IRS, those two beliefs have, at most, *similar* contents.

In general, for IRS, unless there is *identity* of inferential role there is, at best, *similarity* of content. But no two cognitive states have *identical* inferential roles. So there can be no two cognitive states that have the same content: there can, at most, be similarity of content. So two people never grasp the very *same* proposition: thus there is nothing that two people agree or disagree about, for there is no commonly grasped content. But this is surely false. Surely many different people can grasp the same $concept_O$ and the same proposition.

In connection with this, IRS seems to be committed to an untenable form of holism. Let b1 be an arbitrary belief had by Smith, let p1 be the proposition which is believed.; Let b2 be an arbitrary belief had by Jones; and let p2 be the proposition thereby believed. Any two differences in what Smith and Jones believe will be reflected in a difference in the inferential roles had by b1 and b2. There are two reasons for this. Suppose Smith believes *I am in Paris right now* whereas Jones does not. In that case, Smith will be prepared to infer the conjunction *I am in Paris right now and P* whereas Jones will not be prepared to make that inference. So *any* difference in what is believed can readily be reflected in a difference in inferential role. So if IRS is right , any difference in what two people believe (or in what a given person believes at two different times) amounts to a *total* difference in what is believed and, more generally, what is grasped. So unless two people

are in *total* agreement about everything, then there will be *nothing* that they grasp in common. So IRS makes it totally impossible for two people even to *grasp* the same proposition: there can be no agreement or disagreement about anything.

But this is absurd, as previously noted.

There is another related problem with IRS – sometimes it is known as the "error-objection" to IRS. If the inferential role of a concept *determines* its content, then inferences involving that concept cannot be *wrong*: since they *define* the concept, and give it its content, they cannot be wrong *relative* to the content of that concept. Imagine the following. There is some function F such that you have the following beliefs about it. F(4)=6, F(8)=10, and so on: for each x less than a thousand, you believe that F(x)=x+2. But you also believe that F(1000)=1004, and that F(1006)=1010, and so on. If IRS is right, then the fact that you take F(1000) to equal 1004 means that the function you were grasping *all along* was something other than F(x)=x+2. You were grasping some other function – perhaps *for all x less than a thousand, F(x)=x+2, but for all x greater than a 1000, F(x)=x+4*. The point is: if IRS is right, you cannot *wrongly* apply a function that you grasp; for the beliefs you have *determine* which function you are grasping. So IRS makes it impossible for to *wrongly* apply a concept: wrong applications of a given concept ends up being a correct application of some other concept. If you decide that F(1000)=1004, you are not wrongly apply F(x)=x+2: you are rightly apply some "bent" cousin of that function. If IRS is right, then given any concept$_O$ C, an exactly similar argument shows that wrong applications of C collapse into correct applications of some other concept$_O$ C*. (If, with Frege, we think of concepts as being identical with functions – functions from individuals to truth-values – then it is very clear how generalize the argument just made.)

But plainly there is such a thing as *wrongly* applying a function. There is

such a thing as wrongly adding two numbers: it is not as though would-be wrong applications of the addition-function collapse into correct applications of the addition* function, where addition* is some bent version of addition. So IRS is wrong.

Evaluating Fodor's 3rd Argument

I think that Fodor's 3rd argument collapses when we apply an analogue of Kripke's distinction between reference-fixing and meaning-giving.

Let us start with the basics. When we say that somebody has a "concept" of some thing x, what we do we mean? We mean, I think, that that person can think about x – that he can entertain thoughts (of some kind or other) about that thing. To have a concept of Socrates is to be able to have thoughts about Socrates – thoughts like *Socrates was bald.* If we say that Sam has a concept of the number two, we are saying, nearly enough, that Sam can think true or false things about that number. It seems self-contradictory to say: "Sam has a concept of the number two. But he just can't grasp truths of falsehoods concerning that number."

Being able to think about a thing seems to involve being able to *single it out*. To be able to think about the number two, you need to be able to cognitively lock onto it, to the exclusion of all other things.

Consider your concept of the number two. It seems inconceivable (to me, anyway) that somebody would have a concept of that number but *not* know that two is one more than one and one less than three. It seems inconceivable that a person would be capable of grasping the proposition *Fred has two apples* without also grasping the proposition *Fred has more than apple* and *Fred less than three apples*, and so on.

This suggests that certain pieces of knowledge are *constitutive* of one's

concept of the number two. One couldn't possibly believe n=2 if one didn't believe *n=1+1* and *n=3-1*. This suggests that for one to believe *n=2* just *is* to believe *n=1+1*, and so forth.

Fodor denies this because he thinks it leads to an untenable holism. But it is obviously something we *don't* want to give up lightly – and I am now trying to show that it does not lead to an untenable holism.

But there is an incoherence in what we've just said. One cannot have true beliefs *about* x unless one *already* has a concept of x. You cannot believe that two is greater than one or less than three unless you *already* have a concept of the number.

So we have to find a way to accommodate two opposing sets of intuitions. First, it seems inconceivable that one would be able to grasp propositions like *n=2* and *John has two apples* unless one could grasp related propositions (e.g. *n is more than one*, *John has more than one apple*). This suggest that having a concept of the number two involves having knowledge of certain rudimentary facts about that number. And this, in its turn, suggests that one's concept of the number is *constituted* (at least in part) by knowledge of such truths.

At the same time, knowing a true proposition *about* the number two involves *already* having a grasp of that number.

There is an easy way out of all this. The appearance of contradiction is generated by a false premise. Consider the proposition

(*) *somebody x was uniquely a great philosopher to die of hemlock poisoning.*

(*) *describes* Socrates. But to grasp (*) it isn't necessary to *already* have a concept of a Socrates.

Let me a little more precise about what the term "describe" means in this context. (*) is an existential generalization, not a singular proposition. But a

generalization is always made true by some *singular* fact. It isn't possible just for *somebody* or other to cough. Some *specific* person must do so. It must be *Smith* who coughs – or Jones or Brown. So the proposition *somebody coughs* is "made true" by the singular fact that Smith coughs. That particular proposition is made true by many different singular facts – by Brown's coughing, Jones's coughing, and so no. Given a proposition of the form *for some x, x has phi*, let us say that O is a "truth-maker" of that proposition iff *O has phi* is true. So Brown is a truth-maker of the proposition *for some x, x coughs* because *Brown coughs is true.*

An existential generalization of the form *something x uniquely has phi* necessarily has at most one truth-maker. Let us say that a proposition having that form "describes" O iff *O uniquely has phi* is true. Clearly (*) "describes" Socrates (in our technical sense of the word): Socrates is its unique truth-maker.

Now if you grasp a proposition that describes some thing x, then you are in a position to have further thoughts about x – you are in a position to (truly or falsely) attribute properties to x. Suppose Smith thinks to himself: *somebody x was uniquely a great philosopher to die of hemlock poisoning.* Smith can then think: *and that guy probably had a most unpleasant death*. In fact, for any $\text{concept}_{\text{O}}$ phi that Smith grasps, Smith can think *that guy had phi*. Because Smith knows the truth of (*), Smith is now in a position to attribute any property he wishes to the truth-maker of (*). In a word, Smith can think anything he wants about Socrates.

You are talking with your friend Smith, whom you know to be a reliable information. Smith says to you: "Somebody just discovered a cure for AIDS." You don't know *who* discovered the cure. But you are still able to say things like "well, *that* person will be famous", "*that* person has done humankind an immeasurable service", "is *that* person working for a government agency or

for a private organization?", "I bet *that* guy has a thing or two to say about current government policy regarding stem-cell research". Basically, given knowledge of a proposition that *describes* (in our technical sense) someone or something x, you are able to think things *about* x.

What we just said about Socrates and the (future) discoverer of a cure for AIDS can be applied to anything. Consider the number two. Suppose you know the truth of the following proposition:

(**) There is some number n such that n=1+1.

If you know (**), then you are in a position think things about *that number.* You can think "I wonder if Smith has *that number* of cars" and "*that number* is greater than one" and "any closed figured has more than *that number* of sides (if its sides are straight)."

Knowledge of a proposition that *describes* a thing apparently suffices for ability to have thoughts about that thing – so knowledge of such a proposition seems to qualify as a *concept* of that thing. And knowledge of such a proposition does not itself presuppose a concept of that thing.

Given all of this, here is what I would propose. To have a concept of a thing is to know some proposition that describes x (in our technical sense of "describes"). Now many will respond by saying that this view has been straightforwardly refuted by Kripke, Putnam, and others. Such a response, I will argue, involves a number of confusions – mainly, a confusion of linguistic meaning with cognitive content. But before I deal with that set of issues, allow me to develop our theory, and also to show why, if correct, it falsifies Fodor's slippery-slope argument.

Suppose you know the truth of (**). First of all, as we've seen, it

immediately follows that you are able to have thoughts about the thing which (**) describes. (**) describes some number n. If you know (**), then you are able to think *n is larger than one* and *Smith has n cars*, and so forth. Second, if you know (**), then you *automatically* know a certain fundamental truths about n – you know that n=1+1, that n is greater than zero, and so on.

Now we can close the argument. Suppose that, as we are maintaining, your knowledge of (**) is your *concept* of the number two. Let C be your belief in that proposition. In that case, it doesn't matter what else you come to believe (truly or falsely) about the number two. Obviously *C* is not, on that account, going to undergo any change; so your concept of that number isn't going to change. It is true, of course, that the inferential role of that concept will be different: now, given the proposition *Fred has two cars*, you will be prepared to infer that *Bob and Fred have the same number of cars*, whereas, previously, you were not prepared to draw that inference. But none of that is going to change your *concept* of the number two in the slightest. After all, that *concept* is identical with your belief in (**) (or some other proposition that describes two), and that belief isn't going to be changed with changes in what you believe *about* that number.

This analysis has another virtue – it sidesteps the "error-objection" to IRS. On our view, not *every* one of the beliefs you have about the number two is definitive of your concept of that number. Only your belief in (**) is thus definitive. But once you have a belief in (**), you can have any number of wrong beliefs about that number. On our view, to think that the number two has phi is to think something like the following: *for some number n, n=1+1, and n has phi.* Now suppose that the number two does *not* have phi. On our view, you are still having a wrong thought *about* the number two. Suppose Brown has three cars. Given this, let phi be the property of being equinumerous with the number of cars Brown has. On our view, to think *2=the*

number of cars Brown has is to think something like:

(***) *for some number n, n=1+1, and Brown has n cars.*

(***) is a wrong belief *about the number two.* Your wrong belief about two doesn't collapse into a correct belief about three. What you've done is to attribute to Brown the property of having n cars, where n=1+1. So what you've done is to *wrongly* think that Brown has *two* cars. So you've had a wrong belief about the number two: your mis-application of the concept *two* didn't collapse into a correct application of some other concept.

Our analysis has one virtue. The objects of thought are not objects but *propositions.* You don't think *Socrates.* You think *Socrates is bald* or *Socrates was wise.* It isn't possible to think a *thing* – only a truth or falsity about a thing. You don't think *two* – you think *two is even* or *1+1=2.* In any case, this seems plausible –luminaries ranging from Kant to Frege to Brandom have held it – and it is worth seeing how far we can go with it.

A concept is just *is* the ability to grasp a thing. But, as we've just noted, grasping a thing consists in grasping a proposition *about* that thing. So it is hard to see how one could exercise a concept of a thing without thinking a truth or falsity *about* that thing. Our analysis is obviously consistent with this: to have a concept of the number just *is* to know a certain kind of proposition about it.

The other side of the coin is this. Suppose you say that having a concept of (say) the number two does *not* consist in knowing some proposition. In that case, it seems to follow that you could *exercise* your concept of that number – you could *think* about the number two – without thinking anything *about it.* Permit me to clarify the logic involved. To have a concept of (say) the number two is to have a grasp of it. But what is grasped is never a thing *simpliciter*, but

only some proposition concerning that thing. So a concept of a thing must be a concept of a proposition *about* that thing. Otherwise we are saying that one can grasp a thing *without* grasping any proposition concerning it: we are saying that one can *just* think *two*, as opposed to *1+1=2* or *two is even.* But, on the face of it, it very much seems that whatever is thought – whatever is grasped – is some truth or falsity, and never just a bare object. In any case, our analysis is on the right side of the view – which seems reasonable to many, though it will not seem so to many others – that it is whole propositions that are thought, and never just bare things.

So far we have (tendentiously) been speaking about numbers. Let us apply our analysis to less sterile domains, beginning with a story. You don't know who invented bifocals. (It was Benjamin Franklin, as it happens.) Somebody whom you know to be a reliable source of information says to you: "Some one person invented bifocals." On this basis, you are able to have thoughts *about* that person. You can say things like: "Well, I bet *that* person was very proud of his invention" and you can ask things like "did *that* person have poor eyesight? Did that person personally benefit from his discovery?" What is going here seems to be this. You know some true claim of the form: something x uniquely has phi. In this case, the proposition in question is:

(IBF) some man uniquely invented bifocals.

That proposition describes (in our technical sense) some man – namely, Benjamin Franklin. Because you know this proposition, you are able to affirm things, and ask things, and so forth, *about* that person. So your knowledge of (IBF) gives you the ability to think about that person – it is a *concept* of that person.

When you *use* your knowledge of (IBF) to attribute some property phi to that

person, you are really thinking some existence claim: *some man uniquely invented bifocals, and that person had phi*. So when you use that concept to attribute cleverness to that person, you do so by way of the thought: *some man uniquely invented bifocals, and that person was clever.*

Here we must be very explicit and cautious if we are to stay on the right side of some important semantic and modal points. Your concept of the person who invented bifocals is constituted by your knowledge of proposition that describes that person. You grasp him *through* the proposition *unique inventor of bifocals*. But this doesn't corrupt your modal intuitions in the slightest. You know that *that person*, whoever it was, didn't *necessarily* invent bifocals – that he might have died at the age of three or might have decided to choose a path in life that didn't involve inventing bifocals.

Suppose I say to you: "some guy stole my wallet" (i.e. some man x uniquely stole my wallet). For analogues of the reasons just given, you are able to have thoughts *about* that person, whoever it is. You can think "he must be a person of low morality" or "he probably had a bad childhood" or "he is probably a drug addict". So even though your concept of this person is constituted by your knowledge of the existence claim *somebody x uniquely stole __'s wallet*, you still know that that person, whoever it was, didn't *have* to become a thief – that he might have taken a different path in life. He might have become a monk or a university professor. So the fact that, in at least some cases, we know things through existence claims which describe them does not force us to have the wrong intuitions about modality or analyticity.[414]

Let us take stock. First, as we saw, your knowledge of (IBF) constitutes a concept of Franklin. Second, that concept is obviously not an "atom". It presupposes a grasp of other concepts – the concept of bifocals, of being an inventor, of being a male, and so forth. (Of course, possession of any given one of these conceptso presupposes possession of others: consider what you

must *already* grasp to grasp the concepto of bifocals or of being an inventor.) The third point is the most important. Now that you know (IBF), and thus have a concept of a certain person, it doesn't matter what you *henceforth* think about that person: it won't alter your concept of him. Suppose you learn that was also witty: that won't change the concept in question: the concept is constituted by your knowledge of (IBF) – it isn't going to turn into a new concept, or otherwise change, when you learn new things about him. When you learn that that person was one of the Founding Fathers, your *concept* of that person isn't going to change. It is true that, under that circumstance, you will accept some inference that you didn't previously accept: you will accept the inference from *x invented bifocals* to *x was a Founding Father.* But this fact doesn't mean that your concept of that person suddenly changed: that concept remains what it always has been – a knowledge of an existence claim. By the same token, you can come to believe any number of *wrong* things about that person: you can come to believe that he was over seven feet tall or that he invented the airplane. Those beliefs will be wrong beliefs *about* Franklin, not about some other person or thing (and not about *no* thing either). Remember what we said a moment ago: when you exploit your knowledge of (IBF) to attribute a property phi to the person it describes, you are really thinking an existence claim: *some man uniquely invented bifocals, and that person had phi*. So suppose you use your knowledge of (IBF) attribute the property of being over sever ft tall to that person. You do this by thinking the following existence claim:

(IBF7) *some man uniquely invented bifocals, and that person was over seven ft tall.*

In thinking (IBF7), you are attributing a property to the inventor of bifocals

that that person didn't actually have. You are having a *wrong* thought about that person. But this doesn't threaten the integrity of your concept of that person. Roughly speaking, the concept is the left half of (IBF7); the false belief is the right half. You can accumulate any number of false beliefs about the inventor of bifocals, and that won't corrupt your concept of the man: put loosely, the concept stays one side of the comma, and the subsequently acquired wrong beliefs remain safely sequestered on the other side of it.

So here we have a concept of somebody that is *not* an atom and that is not vulnerable to any of the arguments that Fodor (or anyone else) directs against IRS.

Now our analysis of conception does seem to involve *a* kind of holism. As we said, to grasp IBF, you must obviously grasp various concepts$_O$. First, you must grasp the concepts$_O$ immediately implicated in that proposition – *bifocals*, *inventor*, and so on. But, as we noted, a grasp of any one these concepts$_O$ presupposes a grasp of others. To grasp the concept$_O$ of an inventor, you must (I would think) grasp the concepts$_O$ of causality, agency, intentionality (in both senses of the word). To grasp the concept$_O$ of bifocals, you must grasp the concept$_O$ of eyewear, of bad vision, and therefore of vision itself; and a grasp of the concept$_O$ of vision involves the concepts$_O$ of space, time, causality (again). So the ability to grasp (IBF) – the ability to have that sort of concept of Franklin – is by no means innocent: it presupposes a background of conceptual weaponry. But it does not follow, and it is not the case, that our analysis of conception is guilty of an untenable form of *belief* holism.

§ If our analysis is to be at all compelling, we must deal with a major issue. We have argued that to be able to think about an object O just *is* to know some true proposition of the form *something x uniquely has phi*, where phi is

a property had by O. We have argued, in effect, that to have a concept of O is to have knowledge of a *uniquely individuating description* of O.

But in light of some points made by Kripke, it is generally taken for granted that this is simply false. Thanks to Kripke, the following is generally taken for granted: one can think about O *without* having such knowledge; conception is *not* descriptive in nature; in many cases, the descriptive information one associates with an object is incomplete, and thus applies to many things besides that object, or doesn't even apply to that object at all. So it would seem that our "descriptivist" analysis of conception is on the wrong side of some very powerful and widely accepted views.

The viewpoint just described has its basis in two points made by Kripke. The first is a narrowly semantic point – a point about the literal meaning of proper names (expressions like "Socrates", "Benjamin Franklin", and the like). The other point is epistemological.

§ Let us start with the semantic point. It is a datum that "Socrates" refers to Socrates. Kripke (1972) made a strong case that proper names are not synonymous with definite descriptions. Contrary to what was previously thought, "Socrates" is not synonymous with "the great philosopher who drank hemlock" (or any other definite description).

This strongly suggests that "Socrates" does *not* refer to Socrates by way of any description, and that it refers to him *directly.* Thus "Socrates" is *directly referential*: it is a mere label for a person, not a description that singles out that person. From this, it seems to follow that to *understand* "Socrates" is to know what it refers to: nothing more, nothing less. To understand that expression is to associate it with the right *individual*, not some description or Fregean sense.

Any English speaker can *refer* to Socrates by means of the word

"Socrates". Presumably, if a person can understand sentence like "Socrates was wise", then that person is able to grasp what is meant by that sentence. And that, in its turn, would involve grasping the things meant by the components of that sentence – in particular, by "Socrates". For reasons stated a moment ago, it seems that the thing meant by "Socrates" is simply Socrates – it isn't a description or Fregean sense that singles out Socrates.

This brings us the Kripke's epistemological point. Kripke argued[415] that people don't typically know "uniquely individuating descriptions" of the objects to which they refer and about which they think. Any English speaker can *understand* the sentence "Socrates is wise". But if asked to produce a description that uniquely applied to Socrates, nine out of ten people couldn't do so. They would be able to produce *some* description, but not that one applied to Socrates to the exclusion of everyone else, and perhaps not even one that applied to him at all. They might say "he was the author of the *Metaphysics*" (false) or "he was a famous philosopher of antiquity" (incomplete, applies to people besides Socrates)

This seems to show that knowledge of a description that singles out x is *not* necessary for thinking about x. Smith may be able to think about Socrates and refer to him; but the only *description* Smith can associate with Socrates is some faulty or overly inclusive one ("author of the Metaphysics", "philosopher of yesteryear", "philosopher who died of unnatural causes").

§ Let me begin by identifying my position – this will involve stating my biases. First, I myself *agree* wholeheartedly with Kripke's semantic point. "Socrates" is *not* synonymous with "the great philosopher who drank hemlock" or any other definite description. Kripke's arguments for this are extremely compelling. In connection with this, I also *agree* that "Socrates" is directly referential: at the level of semantics, "Socrates" is simply a label for

Socrates. So the proposition meant, at the level of literal meaning, by "Socrates was bald" is simply *Socrates was bald*: it isn't *somebody x was uniquely a great philosopher who drank hemlock, and x was bald* or anything of the sort.

Here is where I *disagree* with the story told in the last few pages. First of all, literal meaning is not epistemology. We must be very careful to distinguish what is *literally* meant by an expression, on the one hand, from what it communicates to us. I am not alluding to the kinds of non-literal meanings discussed by Grice, but to something much more radical. There is a difference between hearing the *sound* "Socrates was wise", on the one hand, and knowing what it means, on the other. One has to compute the meaning of the sounds; and this inevitably involves a great deal of background knowledge. The background information *through* which computes literal meaning is to be distinguished from literal meaning itself. Much of what a sentence-utterance communicates to us is connected with that background information. Kripke is (I believe) exactly right about what is literally meant by "Socrates": that expression is not a definite description. But given only that fact, nothing follows regarding the information *through which* one assigns the right literal meaning to "Socrates was bald", and other sentences containing "Socrates". And, I will argue, the information *through* which one assigns the right literal meaning includes knowledge of the right kind of uniquely individuating description. That description will seldom, if ever, be a historically momentous one, like "the great philosopher who drank hemlock"; it will typically be rooted in the peculiarities of the context in which one first encounters the relevant referring term. But, I will argue, knowledge of a uniquely individuating description is *always* involved in knowledge of what "Socrates" means and what it refers to. Kripke's argument to the contrary has weight only because it focuses on the wrong uniquely individuating

descriptions.

§ The best way to begin is with Kripke's own positive account of reference and conception. In his view, one is able to refer to Socrates, and to cognitively access the man, *not* by virtue of knowing some individuating description, but by virtue of having a certain *causal* connection to the man. [416]

The main idea is that a *causal*, rather than a descriptive, connection to x is what underwrites the ability to refer to, and think about, x. And this idea seems to have a prototype in the fact that to *sense-perceive* an object x is not (merely) to have a mental image that "fits" x, but is to have a mental image that has the right *causal* connection to x. (If you have a mental image with the right phenomenology, but the wrong cause, it isn't a perception of the dog in front of you – it is a hallucination or, perhaps, a perception of some other dog.)

§ What is the theory of reference that Kripke outlines? [417] [418] Let's say I don't know of any uniquely individuating description of Socrates. Still, I can refer to and think about the man. How? Somebody encounters Socrates for the first time. (Presumably this is when Socrates is born, but that doesn't matter in this context.) That person dubs his new acquaintance "Socrates". [419] Socrates goes about, introduces himself. Some people have the good fortune to *personally* see Socrates and be told that he is called "Socrates".

Those people obviously can refer to Socrates. And they can *transmit* that ability to others. They can say: *I knew this fellow named "Socrates". He was the most obstreperous person I ever met.*…Having heard (or even just overheard) this, the auditor of that statement can *himself* refer to, and think,

about Socrates. This kind of transmission can go on ad infinitum. 2,500 years after Socrates' death, we can refer to him, thanks to a long chain of such transmissions.

§ Let us now critique this picture, beginning with a purely terminological point. In this context, I will talk a great deal about "descriptions". Sometimes we use that word to denote certain kinds of linguistic entities, e.g. "the man who stole my wallet", "a great general of antiquity". In this context, I will use that expression to refer to the information *expressed* by such linguistic entities; so I will use the term "description" to refer to the concepts expressed by things like "the greatest French soldier" and the like.

Now to matters of substance. Suppose that your companion points someone out to you and says: "that guy is Fred". As a result, you learn who "Fred" names. Here it *seems* as though you learn who "Fred" names *without* knowing any true proposition of the form: *something x uniquely has phi and "Fred" names x.*

But that is not the case. Whenever an object is successfully indicated to you, you *do* have knowledge of a uniquely individuating description of it. If your friend's ostensive definition of "Fred" is to work, you must be able to single out the thing ostended. He cannot just be *a* guy. He must be some specific guy. And he is – the right kind of uniquely individuating information is easy to come by. He is the unique guy standing *over there*, next to *that* water fountain. Or he is the unique object of your attention at *this* moment in time. Or he is the unique guy now running away from you with your wallet.

The uniquely individuating description through which you lock onto Fred may be, and probably is, highly context-specific -- replete with the kinds of things that, in language, are expressed by demonstratives (*that* guy over *there*…). But here what is important is that the lock-on involves knowledge

of *some* description that singles Fred out: the exact nature of this description is irrelevant.

Let us sum up this leg of the argument. Your friend says "that [pointing to someone] is Fred". If you are to assign "Fred" to the right person, you must have some way of cognitively singling that person out. Doing this involves knowing some description that Fred alone satisfies. That description may be, and probably is, highly context-specific and thus of no historical moment (*the unique guy who just bumped into my car and then drove away without giving me his insurance information…*) But that is irrelevant in this context – what is relevant is you know of *some* uniquely individuating description.

There is more to say. An ostensive definition typically involves an indexical –something that, in English, usually involves a "this" or a "that". Now an indexical always picks out the thing satisfying a certain description. "That man" picks out the thing satisfying the description *contextually salient male* or perhaps *ostended male.* "Today" picks out the thing satisfying the description *day of utterance.* "I" picks out the thing satisfying the description *speaker of utterance.* The same holds (mutatis mutandis) for any other indexical. So order for an indexical to work – in order for it to direct the auditor to the right thing -- the thing thereby referred to has to satisfy some existence claim: it must be a thing x such that x uniquely satisfies a certain description. (Let IDL be an arbitrary indexical-token.) And the auditor assigns the *right* referent to IDL only in so far as he knows some true claim of the form *there is some x such that x uniquely satisfies description D, and IDL refers to x.* So even *ostension,* and thus ostensive definitions, operate through the medium of uniquely individuating descriptions: and assigning the right referent to the ostension involves knowing the right such description.

§ Let us now reconsider Kripke's story in light of these platitudes.

Imagine that there people who personally *saw* Socrates, and were told: *that* guy is named "Socrates". (To simplify discussion, let us naively suppose that the word they heard really was "Socrates".) Each of these people knows some true proposition of the form: *somebody x uniquely has phi and "Socrates" names x.* For the reasons considered a moment ago, each one of these privileged few learns the truth of a proposition like: "Socrates" denotes *the unique person **over there**, next to the hearth, talking about the One, wearing a purple toga…*

Some of these people might actually meet Socrates. In that case, once again, one has a uniquely individuating description. Each of these people learns the semantics of "Socrates" through a description like: "Socrates" denotes *the person who **I** shook hands with at time t, place p, who irritated **me** by contesting my definition of piety…* (The description has reference-fixing, not meaning giving, role.)

Suppose that Larry is one of the people who where given an *ostensive* definition of "Socrates". So Larry *saw* Socrates, possibly met him, and was told: ***that*** guy, over there…is named "Socrates".

Larry's uniquely individuating description almost certainly has an *indexical* component (the indexicals are in boldface). Larry is told that "Socrates" refers to ***that*** guy. Larry *sees* (or otherwise sense-perceives) Socrates: So Socrates is given to Larry through information like (speaking from Larry's perspective): *the guy who **I**'m talking to right **now*** or *the guy ten feet from **me**...*

But that doesn't in any way mean that Larry doesn't have a uniquely individuating description of Socrates. He *does* have such a description. Many uniquely individuating descriptions contain an indexical component.[420]

Let us continue the story. Larry outlives Socrates. Talking to his buddy Fred, Larry says (in Ancient Greek): *I knew this guy named "Socrates"; that*

guy did such and such…

Fred knows a true proposition of the form: *something x uniquely has phi, and "Socrates" names x.* In this case, phi=the property of being uniquely referred to by Fred on occasion C. So on the basis solely of this meager bit of reportage, Fred has a uniquely individuating description of "Socrates", namely: *the guy to whom Larry was referring on occasion C (at place p, time t)…*

Here a couple of points are in order. First of all, there is no circularity in this account. Larry had a uniquely individuating description of Socrates. That is beyond doubt. Now Larry's description was not circular: it did not have the form: "Socrates" names the guy called "Socrates". Larry's uniquely individuating description (uniquely individuating description) was something like: "Socrates" refers to *that* guy – the one next to the hearth in the purple toga…

Now Fred's uniquely individuating description of Socrates isn't circular either. Fred's uniquely individuating description is something like: "Socrates" denotes *the guy to whom Larry was referring on such and such occasion.*…So Fred's concept of Socrates consists in Fred's knowledge of the proposition: *there is some x such that, on such and such occasion, Larry was referring to x, and "Socrates" names x.*) Fred's uniquely individuating description is parasitic on Larry's. But Larry's description isn't circular, and neither is Fred's.

Another point must be made. Larry could introduce Socrates through a highly impoverished description: *But even so, Larry has a uniquely individuating description of Socrates.* This point is crucial and, so far as I know, has always been overlooked.

Suppose Larry says: *once I met this chap named "Socrates". He was a real jerk. I wish I'd never met him. I'm sorry I brought him up. Let's not talk*

about him anymore.

It might seem at first that, under the circumstances, Fred doesn't have a uniquely individuating description of Socrates – that he has only an overly inclusive description: *some jerk who Larry met.*

But that is not so. Fred *does* have a uniquely individuating description : *the guy who Larry was referring to at place P, time T*…For there was *one* guy who to whom Larry was referring at that exact instance – at the specific time when Fred first heard the name "Socrates". So the description through which the semantics of "Socrates" is given to Fred (in the reference-fixing, not the meaning-giving, sense) *is* uniquely individuating and accurate, namely: *the guy who Larry was referring to at place p, time t.…*

Further, there is no way that the ability to refer to Socrates could be transferred from Larry to Fred (or from *any* person to any other) *except* by way of a uniquely individuating description (I mean a description that *in fact* singles out the relevant object). Transference of that power, from Larry to Fred, involves, at the very least, Larry's saying something about Socrates to Fred, or in Fred's presence. And no matter how cursory that something is, Fred *ipso facto* has a uniquely individuating description of Socrates. Suppose that, in Fred's presence, Larry utters:

(*) "that Socrates was a real nuisance. I wish he'd just stayed in Athens."

Given that (*) is Fred's only source of information about Socrates, it might *seem* as though Fred doesn't even come close to having knowledge of a uniquely individuating description of that philosopher: all Fred has is the incomplete, and possibly false, description "a real nuisance". But this is not so. Fred knows that, on such and such occasion, there was some x such that Larry was referring to x and also that "Socrates" names x. So given only the

meager bit of information embodied in (*), Fred *does* know some true proposition of the form *something x uniquely has phi and "Socrates" names x.* In this case, the proposition in question is: *there is some x such that, on such and such occasion, Larry was referring to x, and "Socrates" names x.* [421]

Of course, we can imagine Fred transmitting the powers in question to some third party – say, William. By the logic we just discussed, William's acquiring these powers involves his (William's) acquiring a uniquely individuating description of Socrates. And so on, until the chain we've described reaches you, or your grandchildren.

It must be pointed out that the only *correct* uniquely individuating description that Fred has may be one like: *the guy to whom who Larry referred on such and such occasion.*…Beyond that, Fred's beliefs about Socrates may be quite spurious. Fred may believe that Socrates was a person of low morality, and that he wrote the *Posterior Analytics*, and so on. By the same logic, the only uniquely individuating description one of us – twenty-five hundred years later – has of Socrates may be one like: *the guy who Frank was referring to that time when.*…Beyond that, one's beliefs about Socrates may be spurious.

§ Let us sum up. There are different ways that one can come to be a competent user of a referring term like "Socrates". Such competence can be acquired through an ostensive definition ("*that* guy is named 'Socrates'") or through a descriptive definition (" 'Socrates' is the name of the great philosopher who died of hemlock poisoning". It may acquired on the basis of some threadbare bit of (possibly erroneous) reportage, like (*). But no matter which of these ways is operative, one's competence to use that word *does* embody knowledge of a true proposition of the form: *something x uniquely*

has phi, and W names x, where W is the referring term in question.

This is not in tension with Kripke's semantic insights. Kripke is probably right to say that "Socrates" is not synonymous with a description. But semantics is not epistemology. Suppose, for the sake of argument, that the semantic rule for "Socrates" is some bare pairing like *"Socrates" refers to Socrates*. To grasp that rule, you have to be able to pair off "Socrates" with the right person. So you have to have some way of thinking about that person. Socrates is not given to you in a vacuum. He is given to you through verbal communications or through your own perceptions. So your epistemic *access* to Socrates, and therefore the semantic rule just discussed, embodies knowledge of the right kind of uniquely individuating description. But the *semantic rule* itself is not, at least not no that account, descriptive: it may be a bare pairing.

§ So embodied in one's competence to use an expression like "Socrates" *is* a knowledge of a uniquely individuating description that singles out the referent. But that uniquely individuating description may be highly specific to one's own circumstances, and may thus be of no interest to historians or anyone else who was interested in Socrates' biography. Russell – *rightly*, as I think we are seeing – thought that to refer to Socrates one had to have a uniquely individuating description of the man.[422] But, judging by his examples[423], he *wrongly* took the uniquely individuating descriptions in question to be historically momentous ones: *the guy who invented the theory of forms and drank hemlock.* He was wrong about the nature of those uniquely individuating description, but right, apparently, that one must have *some* uniquely individuating description.[424]

§ Our account is *consistent* with the idea that one's ability to refer to Socrates involves a *causal* connection between Socrates and oneself. The account we've given *is* a causal account, even though it is also a descriptivist account. Larry *sees* Socrates: obviously there is a causal connection there. And that is how Larry gets *his* uniquely individuating description of Socrates. Fred sees or hears Larry refer to Socrates on some particular occasion: obviously there is a causal connection there. And so on, until this chain reaches you. So you *are* causally connected to Socrates: and it *is* this causal connection that enables you to refer to and think about Socrates. But this causal connection is embedded in various uniquely individuating descriptions (or perhaps it would be better to say that these uniquely individuating descriptions are embedded in it).

Clearly not just *any* causal connection between a person and a thing enables that person to refer to that thing. The causal connection operates *by way of* some conception. I am directly causally connected to many individual atoms (which bounce off many surfaces). But given some one of those atoms – call it "Alpha" – can I refer to Alpha? Is my being hit by Alpha *enough* for me to refer to it? It seems not. I have to have some way of *singling out* Alpha in thought. That way of singling out Alpha may – indeed probably will – exploit my causal connection to Alpha. But it is not the causal connection *by itself* that enables me to think about or refer to Alpha. What gives me that ability is some conception which *exploits* that causal connection.

§ In a number of publications, Putnam has argued for a "linguistic division of labor". It might be thought that, given what Putnam says in this connection, we clearly *can* refer to things without knowing descriptions which pick them out.

The idea would be this. Some *scientist* has knowledge of the elaborate body

of theory underlying the idea that (say) there are electrons. On this basis, the scientist is able to think about, and refer to, such things. The ordinary non-scientist does not have the just mentioned theoretical knowledge. But he is still able to think about, and refer to, such things: his ability to do so is parasitic on the scientist's ability to do so. So the scientist has the right descriptive-theoretical knowledge; the layperson does not; but thanks to the scientist, the layperson doesn't *need* such knowledge to think about, or refer, electrons.

There surely is a linguistic, and cognitive, division of labor. But it does not follow, and it isn't the case, that one can refer to electrons, or genes, or anything else, without having knowledge of the right kind of uniquely individuating description. Let's suppose that scientist Smith knows everything there is to know about electrons; so he knows some theoretical, uniquely individuating description that applies to all and only electrons. One day Smith is talking to his friend Brown, who doesn't know anything about electrons. Smith says:

(#) "I am really fascinated by electrons; I want to study them for the next fifty years."

At that point, Smith quickly changes the subject.

Given that (#) embodies everything that Brown knows about electrons, it might seem as though Brown does *not* have knowledge of a description that applies to all and only electrons. But this is not so. Brown knows that, on such and such occasion, Smith used the term "electron" to refer to a certain type of object. On the basis of (#), there is some property phi such that Brown *does* know some true proposition of the form: *there some type object X such that X has phi, and members of X are called "electrons"*. In this case,

phi is the property of being referred to Smith on such and such occasion.

Brown doesn't know anything theoretically significant about electrons. He doesn't know how they are different from quarks or atoms. In fact, he doesn't even know if they are sub-atomic particles. But Brown *does* have knowledge of a uniquely individuating description that is made true by the type *electron*.

The kind of cognitive and linguistic division of labor still exists. Brown can't do any scientific heavy-lifting: *his* uniquely individuating description is useless for that. (So here we have a cognitive division of labor.) Moreover, Brown's use of the term "electron" presupposes Smith's knowledge of the relevant theoretical machinery. (So here we have a linguistic division of labor.) But, as we saw, Brown does have knowledge of *a* uniquely individuating description; and, in this context, that alone is the relevant fact.

§ Let us now generalize some of the points just made. There are several directions of generalization.

Suppose that you acquire a concept of Socrates by *seeing* him. Here, it seems, your concept of Socrates does *not* embody knowledge of a proposition of the form *something x uniquely has phi*, where phi is a property that Socrates has. After all, you just *see* Socrates; he isn't given to you through a description or, therefore, through an existence claim.

But we must remember that perception is necessarily *descriptive.* The content of your perception is not (just) *Socrates*; it is a whole state of affairs; you see *a* thing in a certain place, having certain features, standing next to a rock of a certain shape…So what is given to you by the perception is existential: there is *a thing* x such that x (and x alone) is in such and such a place and has such and such features. When you see a thing, you cognitively single it out; you cognitively "lock on" to it; and that singling out is

accomplished by the descriptive information embedded in your perception. Embedded in your perception is descriptive information that *singles out* Socrates.

Of course, your perceiving Socrates involves an ineliminable causal connection with that individual. But that causal connection is embedded *within* knowledge of descriptive content. More precisely, it has its lair in the "idexical" or "egocentric" components of that description. What is given to you in percetoni is (just) Socrates; it is *a* man *over there*, in *that* place…A person with a qualitatively identical perception *could* be seeing a state of affairs in some completely different place. The reason you are seeing a state of affairs in *that* specific place is that you are causally connected a state of affairs in that place: your perception was precipated by photons bouncing off of various things in that specific place. Someone having a perception qualitatively just like yours could be seeing a man in some completely different place. The reason you are seeing a man in *that* specific place – over *there* – is that your perception is precipitated by photons having their origins in a satate of affairs involving a man in *that* place, and not on the other side of the galaxyi.

So, yes, a causal connection between yourself and Socrates is indispensable to the fact that you are seeing Socrates, and not his molecular-twin. But that is only because this causal connection is indispensable to your being given the right kidn of *descriptive* way of singling out Socrates: what is given to you in the perception is *a* thing x that is uniquely in *that place*, and also at *this* particular time, such that x such and such freatures…The causal connection to your environment rivets you to the right state of affairs by way of the egocentric components in the relevant *descriptive* information.

To sum up, when you upload a concept of Socrates by *seeing* him (or, for analogues of the easons just given, by otherwise sense-perceiving him), your

concept of the man *does* embody knowledge of existential information that *singles out* Socrates. What is given to you in the perception is *a* thing uniquely having certain properties – uniquely occuping *that* place, at *this* time…. Since Socrates uniquely has the properties in question, you are seeing *Socrates.* But your seeing Socrates, as we put it, consists in your visually being given a description that *singles out* Socrates: it consists in your being given existential information that "describes" (in our technical sense) Scorates, and Socrates alone.

An exactly similar argument applies to our concepts of natural kinds. Let's say that you *see* a specimen of some natural kind – say water – and, on that basis, upload a concept of that natural kind. So you are able to have thoughts about water; but you don't know that water consists of H_2O molecules. Are we to say that, under this circumstance, you *don't* know of a uniquely individuating description of the form *some type X of object uniquely has phi*, where phi is a uniquely property had by water?

No. Let us consider your perception. Let S be the specific specimen of water that you are seeing. Remember what we said about Socrates. The exact same thing is true of S. Your seeing *S* consists in your being given descriptive information that singles out S. Your causal connection to S is crucial – but only because it is crucial to your being given the right descriptive information. Your seeing S consists in your visually being given a whole scenario:

(*) there is some body of transparent liquid x such that x is uniquely over *there,* in that glass….,

where S is the relevaqnt body of liquid. That existential information uniquely "describes" S, in our technical sense of the word; and that is why you are

seeing S. So your concept of S embodies knowledge of the right kind of uniquely individuating descriptive information.

The same is true of your concept of the corresponding natural kind. By supposition, you acquire a concept of the natural kind *water* by seeing some specimen of that kind: you *see* S. So your concept of water is given to you through the description *the kind of which S is an instance.* But remember that your concept of S, in its turn, consists in your knowledge of a description – namely (*) that describes S. So your concept of water is given by the description:

(**) the kind X such that, there is some body of transparent liquid x such that x is uniquely over *there*, in that glass…and X is the kind of which x is an instance.

So your concept of water is given by your knowledge of an existence claim, namely:

(***) There is some kind X such that, there is some body of transparent liquid x such that x is uniquely over *there*, in that glass…and X is the kind of which x is an instance,

where water is the unique natural kind "described", in our technical sense by (***).

The reason that your perception gives you a concept of *water*, as opposed to twin-water, *is* that you have a causal connection to a specimen of water. But, as we've seen, that causal connection is indispensable only because it is indispensable in giving you the right descriptive information; that causal connection rivest the egocentric components of that description (over *there*…

now.…) to the right place and time: a place and time where the salient state of affairs is one involving water, as opposed to some other substance.

It will be said that:

One's concept of water cannot *possibly* consist in knowledge of some existential information like (***); for no one has *any* recollection of their first perception of water; any such recollection vanished long ago. Perhaps one's concept is *initiated* by such knowledge. But that information soon drops out, leaving one with a pure, object-tracking awareness.

I deal with this sort of consideration at the end of this paper.

Let us sum up what we've seen so far. In some cases, one acquires a concept of a thing by hearing about it – *words* are involved. We've seen that, whenever that is the case, one's concept of that thing involves knowledge of the right kind of existential proposition. In some cases, one acquires a concept of a thing in a more direct way: one *sees* it (or hears it…). We've seen that, in that case, one's concept of that thing embodies knowledge of the sort of proposition just described.

But there is another way of "uploading" a concept of a thing. One can *infer* its existence. This is we learn about so-called theoretical entities and kinds. But, of course, not all things learned of in this way are posited in the course of scientific theorizing. Suppose you find that your house has been burgled. On the basis of the evidence, you realize that there was a single burgler. It seems that you have a acquired a concept of a certain individual through inference; but, of course, that individual is not a "theoretical entity", at least not in the sense in which electrons are.

It is pretty clear that when we acquire a concept of a thing through inference, our concept embodies knowledge of the proposition just described.

Consider the case of the burgler. What you know is this:

(#) there is some x such that x uniquely burgled my house….

On this basis, you can think things about the burglar: *that* guy must be on drugs; *that* guy's father must have been an overly lax discplinarian…You are able to have thoughts about some specific individual because you know (#). It is your knowledge of that proposition that enables you to think about that burglar. So your knowledge of (#) constitues your concept of the burgler.

The same thing (mutatis mutandis) applies to "theoretical" entities. Astronomers had to *infer* the existence of Uranus. Given the behavior of things which they could see, they knew of the existence of something that they couldn't see. What they inferred was not *Uranus* (objects cannot be inferred, only propositions about objects); what they inferred was this:

(##) given that such and such bodies are behaving in such and such a way, there must be some x having such and such properties (having a certain location and mass…).

Astronomers were able to have thoughts about Uranus because they knew some proposition like (##) – some propsitoin which describes it.

The same is true (mutatis mutands) of our concepts of theoretical kinds. Given the fact that there are such and such observable phenomena, we know that there must such and such unobserbable entities. What we know is:

(###) given that there are such and such observable there must be such and such unobserbable bodies, having such and such properties, behaving in such and such a way…

In other words, what we know is something like:

(###)given that there are such and such observable there must be some kind of object X such that there are instances of x,

where (###) describes the natural kind *electron.*

Our knowledge of inferred entities and kinds is *blatantly* descriptive. It is *clear* that our knowledge of such things consist in our knowledge of the right kind of descriptive propostitions.

We have found that, surprisingly, the same is true of non-inferred (directly perceived) entities, and also of entities that we learn about through (verbal) reportage.

It thus seems, where *any* external object x is concerned, one's concept of x embodies knowledge of existential information that "describes" x, in our technical sense of the word. (Whether the same is true of our own mental states is a question I leave aside.) Thus far, we have fouind *no* exceptions to the principle that conception consists in knowledge of the right kind of existence claim.

§ In any case, none of Kripke's (or Putnam's) semantic insights threaten the idea that one's ability to use an expression that *refers* to Socrates embodies knowledge of a true proposition of the form: *something x uniquely has phi, and "Socrates" refers to x.* Knowledge of such a proposition is apparently *presupposed* by the ability to use an expression to refer to Socrates.

§ Now we can return from the steamy jungles of epistemology to the cool waters of semantics. Given what Kripke argued (not to mention Putnam,

Soames, and others), it seems pretty likely that *Hesperus has phi* has for its literal meaning the proposition: *Hesperus has phi.* But nobody will deny that *Hesperus has phi* and *Phosphorous has phi* have (or may have[425]) very different cognitive values. Some have reacted to this latter fact by trying to come up with a new and better form of descriptivism. But given what we said a moment ago, I don't think such a drastic measure need be taken.

We saw that -- notwithstanding what is, or seems to be, alleged in *Naming and Necessity* -- in order to know what is meant by "Socrates", one *does* have to know some existence of the claim of the form: something x uniquely has phi, and "Socrates" refers to x. Now when you hear a sentence-token, you don't *automatically* know its meaning. In principle, one can hear the sound without knowing the meaning. (This would happen if you didn't speak the relevant language or if you had sustained brain-damage of a certain kind.) You must make your way from the sound to the meaning. You can't do that in a vacuum; you have to exploit your background knowledge. As we saw, if you know what (to switch examples) "Hesperus" refers to, it is by way of some existence claim, which might be something like:

(#H) *yesterday there was a unique star x such that I was gazing at x and x alone remained in the morning sky after all the others had disappeared; moreover "Hesperus" names x.*

If you are to understand an utterance that you hear of *Hesperus has phi*, you have to hook up that token of "Hesperus" with the right referent. And to do that, you have to exploit the existence claim that embodies your knowledge of what it refers to. So to assign the right referent to "Hesperus", you have to work through some existence claim like (#): that existence claim (or some other) is your only lead. So even though what *Hesperus has phi*

literally means is simply *Venus has phi*, nonetheless you can assign it the right meaning only via (#). So what "Hesperus is lovely" ends up imparting to *you* is some descriptively rich existential proposition:

(#HL) *yesterday there was a unique star x such that I was gazing at x and x alone remained in the morning sky after all the other stars had disappeared; moreover "Hesperus" names x*; and *x is lovely.*

For exactly similar reasons, what "Phosphorous is lovely" conveys to you will be something like:

(#PL) *yesterday there was a unique star x such that I was gazing at x and x alone appeared in the evening sky before all the other stars; moreover "Phosphorous" names x*; and *x is lovely.*

So even though "Phosphorous is lovely" has for its literal meaning some descriptively impoverished singular proposition (*Venus is lovely*), nonetheless what it conveys to you – thanks to the descriptive information through which you *assign* it the right meaning – is some descriptively rich existence claim. It is clear why, given these principles, "Hesperus=Phosphorous" will communicate a complicated existence claim, even though what it literally means is some bare identity.

§ This analysis is on the right side of Kripke's powerful modal and semantic points. Consider the sentence:

(*) "Hesperus is the last star to disappear from the morning sky."

Kripke makes the (correct) point that (*) is not *analytic*. "Hesperus is *not* the last star to disappear from the morning sky" is not self-contradictory.

Our analysis is consistent with this. Let us suppose, once again, that the proposition through which you assign "Hesperus" to Hesperus (Venus) is this:

(#H) *yesterday there was a unique star x such that I was gazing at x and x alone remained in the morning sky after all the others had disappeared; moreover "Hesperus" names x.*

Notice that "Hesperus" gets *narrow-scope* with respect to the descriptive information. Nothing – no descriptive information, no Fregean sense – comes between *Hesperus names* and the thing named. To use Kripke's important distinction, the descriptive information has a "reference-fixing" but not a "meaning-giving" role. So you know that "Hesperus" is *just* a name, and *not* a description, and that all it does is to label a certain thing. So on our analysis (*) is *not* analytic – at the level of literal meaning, it encodes a proposition of the form *O the last star to disappear from the morning sky.* It does not encode a proposition of the form *the last star to disappear from the morning sky is the last star to disappear from the morning sky.*

Further, on our analysis, you *know* that (*) is not analytic: you have the right intuition about this. You assign "Hesperus" to Hesperus by way of your knowledge of (#H). In that proposition, the descriptive information is given "wide-scope". So you know that no descriptive information comes between "Hesperus" and the thing it labels; you know that "Hesperus" is *just* a label, and not a description. So you know that (*) is not analytic. So you have the right intuitions where the modal status of the *sentence* is concerned.

In connection with this, our analysis is consistent with the fact that you

also have the right intuitions about the modal status of the *proposition* meant by (*) and also about the modal properties of the thing named by “Hesperus”. Suppose that the (#H) is the proposition through which you affix “Hesperus” to right entity. In knowing (#H), you know that “Hesperus” names the celestial body which *in fact* disappears from the morning sky after any other. But you will believe that the thing which “Hesperus” names *had* to have that property. In virtue of knowing (#H), you know that something x *in fact* is a celestial body that disappears after all others from the morning sky. Your knowledge of (#H) is not knowledge to the effect that something *had* to have that property. That knowledge is consistent with the position that the thing which *in fact* had that property might not have had it at all. For any object x having that property, your knowledge of (#H) is consistent with your believing the truth of the proposition *x might not have been a last celestial body to disappear from the morning sky.* Put in positive terms, for any object x, your believing (#H) is consistent with your believing the proposition *x might not have been a last celestial body to disappear from the morning sky.*

Another example might be appropriate. Suppose you believe the proposition

(**) “Benjamin Franklin” names the unique person x who *in fact* invented bifocals.

There is no object x such that, in virtue of believing (**) to be true, you believe that x *had* to invent bifocals. Put in positive terms, for any object x, your believing (**) is consistent with your believing the proposition *x did not have to invent bifocals – x might have done something else with his life.* So our descriptivist account of conception does not saddle anyone with the wrong modal intuitions about propositions or objects.

On our descriptivist account, you do indeed think of Hesperus (Venus) by way of some descriptive proposition. But you still have the right modal intuitions. This is because your concept of Venus consists in knowledge of a proposition of the form *something x* in fact *has phi*: it does not consist in knowledge of a proposition of the form *something x* necessarily *has phi.*

One other point should be made explicit. On our account, it is necessary that you think of Hesperus by way of *some* claim of the form *something x uniquely has phi.* But there is no *one* proposition of that form such that you must think of Hesperus by way of that particular proposition. On the contrary, it is essential to our analysis that it be possible to think of Hesperus by way of *different* propositions of that form. That feature of our analysis is essential to our way of solving Frege's puzzle. So our account does not by any means make it necessary that you think of Hesperus *as* the last celestial body to disappear from the morning sky. It allows that you can think of it through infinitely many different descriptions.

That said, it might seem that there is *still* a sense in which our analysis is not modally innocuous. On our account, you think of an object O s *by way of* your knowledge of proposition of the form *something x uniquely has phi*, where phi is a property had by O. So, on our account, you think of Hesperus through *some* such proposition. Of course, our account allows that there are several such propositions. But, for the sake of argument, suppose that in your case, there is only one such proposition; and suppose it is:

(****) *something x is uniquely being a last celestial body to disappear from the morning sky.*

So on our account, when you think *Hesperus has phi*, you are really thinking:

(*****) *something x in fact has the property of uniquely being a last celestial body to disappear from the morning sky,* and x has phi.

So what you think when you think *Hesperus is lovely* is:

something x in fact has the property of uniquely being a last celestial body to disappear from the morning sky, and x is lovely.

A special case of this principle is that, when you think

(A) *Hesperus is the last a last celestial body to disappear from the morning sky,*

what you are thinking is:

(B) *something x is uniquely being a last celestial body to disappear from the morning sky,* and *something y is uniquely being a last celestial body to disappear from the morning sky,* and x=y.

Here is the apparent problem for our analysis. The proposition:

(A_N) Necessarily, *Hesperus is uniquely a last celestial body to disappear from the morning sky*,

is simply false, and nobody thinks otherwise. But what about (B)? At first glance, (B) looks like the kind of proposition that might be necessary. If that is the case, then our analysis is false.

The truth is that our analysis does not put us on the wrong side of any modal facts.

Let us consider the operator *necessarily.* This operator can combine in different ways with (B): more exactly, it can be given varying degrees of scope.

(B1) **Necessarily**: *something x is uniquely a last celestial body to disappear from the morning sky,* and x is identical with a *unique a last celestial body to disappear from the morning sky.*

(B2) *Something x is uniquely a last celestial body to disappear from the morning sky,* and **necessarily**: x is identical with a *unique a last celestial body to disappear from the morning sky.*

(B1) and (B2) are simply false; and nobody would ever think otherwise. So in the relevant respects, they just like (A_N). So our analysis does *not* turn true modal claims into false ones, or vice versa.

Nonetheless, it might seem that our analysis has a peculiar consequence. We have said that you think of Hesperus through either (****) or through some other proposition of the same form. So it seems that, on our analysis, you *must* think of Hesperus a certain way: you *must* think of it as a last celestial body to disappear from the morning sky or a first celestial body to appear in the evening sky: in any case, you must think of it through *some* uniquely individuating description. But this seems wrong: for there doesn't seem to be any that anyone *has* to think of Hesperus. For reasons we've seen, it seems that cognitively "locking onto" a thing O involves knowing some proposition of the form *something x uniquely has phi,* where O has phi. But

after that lock-on has been established, I don't *have* to think about O by way of that proposition. I can drop the proposition. Suppose that I lock onto Sam by having a visual perception of him. In that case, as we saw, I *do* lock onto Sam by way of some proposition of the form of *something x uniquely has phi*. In this case, that proposition might be *somebody x is uniquely a person standing over there, at this time.*…or perhaps *somebody x is uniquely a person whom I am focusing on at this time.*… So forging a concept of a Sam involves knowing some such proposition. But once that concept has been forged, I can, and probably will, forget that proposition. Suppose Sam becomes a life-long acquaintance of mine. I will almost certainly *not* remember the proposition through which I initially locked onto him. So it *seems* that, after a certain point, the uniquely individuating proposition can simply be dropped: after which point, one has a "pure", non-descriptive, object-tracking concept. So, after a certain point, one does *not* have to think of Sam (or anything else) by way of some descriptive proposition. If that is right, then obviously our analysis is false. After all, our analysis nearly enough just *is* the claim that conception is knowledge of some such description.

Here is what I will argue. Any given descriptive proposition can be dropped, and probably will be. But at no point does conception become "pure". What happens is simply that one's knowledge of the old proposition is replaced with knowledge of a new (but related) proposition. At any given point in the career of one's concept of a thing, one's ability to think about that thing consists in one' s knowing *some* such proposition. But there is typically not some *one* such proposition such that, any any point in the career of one's concept of a thing, one must think about that thing through that specific proposition. Let us now turn to this matter.

§ One can easily forget the information *through* which one first acquires a conception of an object. Indeed, this is what happens most of the time. I meet Sam in 1985. At that point, I am told: "that [pointing to some guy on the jungle gyms, wearing a hat with a propeller on it…] is Sam". So it is indeed through some kind of existence claim that I hook up the name to the right person: *there is some guy x who was just pointed out to me, and whom I am now looking it, who is over there on the jungle gyms, wearing a hat with a propeller on it…and "Sam" names x.* But it would be very easy for me to forget that "start-up" existence claim. In fact, with very few exceptions, I *never* remember those start-up claims; they drop out.

This might be taken to show that, at some point, I acquire a "pure" conception of Sam: a conception that is no longer paved with knowledge of existential-descriptive information. So the idea would be this. Perhaps knowledge of a uniquely individuating description was needed to *initiate* the concept. But after that, it isn't needed anymore.

I think this is a big mistake. The "start-up" existential information certainly does drop out. I may not remember, even sub-consciously, the circumstances under which I met Sam. But this absolutely does *not* mean that, at any point, I have some kind of "pure" conception of him, i.e. one that is not paved with existential information. Indeed, if such "pure" conceptions were possible, it is hard to see how it could be possible for Hesperus/Phosphorous cases to arise. But, I fully grant, the point in question definitely requires that we supplement our analysis.

Suppose my first encounter with Sam is through existential information like: *there is guy wearing a green shirt and a bowler hat, whom I am now (on occasion t) meeting and…*and, consequently, that such information is, at least initially, implicated in my interpretetings of "Sam".

I meet Sam again, four days later. And I do so with knowledge: I *recognize*

him. What is happening here? My recognizing him obviously doesn't consist in my becoming aware of some bare identity; it doesn't consist in its being the case, that for some object O, I suddenly become aware that *O is identical with O.* No doubt we do recognize the truth of such bare identities. But that is not what *recognition* consists in. Recognition is not so informationally impoverished.

The story I would tell is this. I realize that there was some chap whom I met a few days earlier, on occasion t, who had such and such features (e.g. he was wearing a bowler hat)…and I also realize that, right now, on occasion t*, there is some chap, who has thus and such properties (e.g. he is wearing a baseball cap)…and I also realize that there is some *one* individual who uniquely satisfies both of these existential propositions.

Let E be the first existence claim, and let E* be the second.

We see that recognizing a thing consists in seeing that some one thing uniquely satisfies each of *two* existence claims. This means that E -- the "start-up" existential information -- can be dropped. My *recognizing* Sam involves my seeing that some one entity satisfies both E and E*. Having made this recognition, I am at liberty to drop E. My awareness of Sam will be sustained, for the time being, by my knowledge of E*. And, in its turn, E* may be dropped and replaced by a third existence claim.

So, I would argue, what sustains a conception of an entity is knowledge of a series of interlocking existence claims. Any given one of these claims may be dropped, and probably will be. But at any juncture in the career of one's having a conception of a thing, that conception involves knowledge of *an* existence claim. And this point, as we've seen, enables us to account (within a Kripke-friendly semantic system) for the cognitive difference between "Hesperus is lovely" and "Phosphorous is lovely". [426]

Intentionalism and Foundationalism

Abstract: Intentionalism is the doctrine that mental states cannot differ in phenomenal content unless they differ in respect of what they represent: phenomenology "supervenes" on representational content. Intentionalism is thus committed to the view – in fact, it *is* the view – that a given kind of phenomenal content *intrinsically* has a certain representational content: nothing must be *added* to that kind of phenomenology to associate it with that kind of representational content; no contextualization, no back-ground knowledge, no ulterior factors are needed to secure that association. (Compare: given that the aesthetic properties of the statue supervene on its physical properties, it follows that nothing must be *added* to those physical properties to secure the connection between those physical properties and those aesthetic properties.) Wilfred Sellars argued compellingly that what makes a given bit of phenomenal content have a given kind of representational content is *precisely* the addition of such background knowledge – of such contextualizing information: to think otherwise, said Sellars, is to be guilty of "the myth of the given" – a spurious epistemic

foundationalism. In this paper, I merely bring out the connection between intentionalism and the myth of the given. I argue that, unless Sellars is completely wrong, intentionalism must be completely wrong.

The purpose of this paper is to show that a doctrine known as *intentionalism* is guilty of what Wilfred Sellars called 'the myth of the given', and is therefore untenable.

What is intentionalism? Some mental entities have both phenomenal content (phenomenology) and representational content. My perception represents the world as being a certain way; so it has representational content. And there is 'something it is like' to have that perception; so it has phenomenal content (it has 'phenomenology').

Intentionalism says that two experiences cannot have different phenomenal contents *unless* they have different representational contents. So phenomenal content 'supervenes' on representational content.

I submit that intentionalism is guilty of the 'myth of the given'. Let me say what this means by way of a Sellarsian story.

You are in a tie shop in which the lighting conditions are a bit unusual. [427] You are looking at a tie in a display case. The tie induces various visual sensations in you. On the basis of these sensations, you can't quite tell whether the tie is orange or red or maroon. You take the tie outside to examine it in better lighting. You see that it is maroon. Now you know what a maroon tie looks like under the lighting conditions that obtain in the tie shop. So the next time you see what is in fact a maroon tie in the tie shop, you *don't* have to take it outside to determine that it is maroon.

But if a tie is some color other than maroon– say, burgundy or maroon --

you cannot know that on the basis of the visual sensations you have in the tie shop – for the same reason that, earlier, you couldn't tell that a tie was *maroon* on the basis of the sensations you had in the tie shop. Let me illustrate my meaning with a new story. You are again in the tie shop, and are looking at what is in fact a burgundy tie. You can't tell what color it is. So you take it outside. There you *see* that it is burgundy. You take it back inside and attend to its appearance. You then know that *that* is what a burgundy neck-tie looks like in the conditions that obtain in the tie shop.

You repeat this process over and over again, until you can tell *for any color C* what an object with C looks like in the lighting conditions that obtain in the tie shop. The correspondence that obtains between objects' colors *in the tie* shop and one's visual sensations is different from the correspondence that obtains between objects' colors and one's sensations outside the tie-shop. By repeating the processing described a moment ago, you learn the former correspondence-rule. And you no longer have to go outside to know what color a tie is. After a while, if you manage to internalize this rule of correspondence, you will, while inside the tie-shop, be able to just *see* what color a tie is. No doubt somebody who has been working in the tie-shop a long time can *see* what colors the ties have; he doesn't have to go through a laborious process of inference (or, at least not a conscious one).

What does this tell us? When you are in the tie shop, your visual sensations do not *by themselves* tell you what color a given object is. What tells you this is your visual sensation *plus* knowledge of various environmental conditions that obtain. By itself, your visual sensation is color-indeterminate. What is color-determinate is the visual sensation *plus* a great deal of background knowledge.

So, when you are in the tie-shop, your visual sensations (at least so far as they relate to color-facts) are not 'foundational'; in order for them to tell you

what colors objects are, they must rely on ulterior knowledge that you have: their ability to tell you anything *presupposes* knowledge, on your part, of various environmental conditions.

Let us sum up. In some contexts, your sensations do not *by themselves* tell you what colors objects are; only those sensations coupled with background knowledge about the conditions of perception tell you that.

It is easy to generalize this argument – to make it apply to *any* perceivable property P. (By a 'perceivable' property, I mean one that we *sense*-perceive objects to have. We don't sense-perceive that an object consists of atoms with three electrons. So that property is not perceivable. But we do sense-perceive that an object is more than ten feet away from us. So that is a perceivable property.[428])

Ordinarily we think that we can just *see* how distant an object is from us. But when I just *see* that a certain tree is less than a hundred feet from me, a lot of background information is at work. For example, you take it for granted that the tree is not 10,000 feet tall. If you thought that the tree *was* 10,000 feet tall, then a phenomenally identical perception would lead you to think that you were seeing a tree was miles away, as opposed to 80 ft away. Let us amplify and substantiate these remarks through a Sellars-type thought experiment.

When you look through a microscope, small things appear big. When you look through a telescope, big things appear small. Imagine that, every morning, you woke up wearing a pair of goggles that either magnified or diminished the strength of your sight. On Monday, you wake up wearing a pair of microscope-goggles – so it would be as though you were looking through a microscope. Another day, you wake up wearing a pair of telescope goggles – so it would be as though you were looking through a telescope. And everyday it would be different; the magnifying power of the goggles

would be different every day. (So one day it would be as though you were looking through a weak magnifying glass. Another day, it would as though you were looking through a moderately powerful telescope.)

Under these circumstances, your visual perceptions at the start of any given day would be distance-indeterminate and, therefore, size-indeterminate. You wouldn't know whether you were seeing an object that was 100 feet tall and a light-year away or an object that was ten feet tall and was ten feet away. You would have to *learn* what correspondence obtained between the character of your visual perceptions, on the one hand, and the distances and sizes of objects.

One way you could do this is as follows. You look at a given object; you don't know whether it is a ten foot object that is 100 ft off or a ten-thousand foot object that is ten miles off. You take off the goggles and you simply *see* how far off, and how tall, the object was; you *see* that it is a building that is approximately 100 ft tall that is 100 ft away. You then put the goggles back on and, now *knowing* the distance and size of the object, you look at it again: you know now that an object that, when you are wearing the goggles, looks like *that* is really an object of such and such dimensions and thus such distance away from you. You repeat this process, with objects of other sizes and distances. And eventually you learn the correspondence that obtains between, on the one hand, the character of the visual perceptions that you have while wearing the goggles and, on the one hand, the distances and sizes of the objects you are seeing. Eventually, you will, while wearing the goggles be able to simply *see* that the object your are looking at is (say) ten feet tall and a 100 ft away – just as somebody who works in the tie-shop can simply *see* that a given tie is burgundy, even though the lighting conditions are non-standard. But, of course, when you simply *see* (while wearing the goggles) that a given object is ten ft tall and a 100 ft away, it is not your visual

sensations *by themselves* that apprise you of the object's size and distance; what really does this is your visual sensations *plus* a great deal of background knowledge about the conditions of perception.

It is important to note that, before you learn the just mentioned correspondence, your visual sensations are distance- and size-indeterminate. A given sensation could tell you *either* that an object is 10,000 ft tall and 100,000 ft away or that it is 20 ft tall and 10 ft. away. Until you acquire the relevant background knowledge, your visual sensations are simply *indeterminate* in these respects; your visual sensations don't tell you *either* that the object is 10,000 ft tall and a 100,000 ft away or 20 ft tall and 10 ft away. In response to this, one might say:

Under these circumstances, your visual sensations *would* give you determinate messages, but these messages would be wrong. You wake up in the morning wearing goggles that, let us suppose, quadruple the strength of your vision. In that case, your visual sensations give you systematically false messages: you *see* objects as being four times larger than they are and, thus, four times closer than they actually. So it isn't that your visual sensations are size- and distance-indeterminate; it is that they *are* size- and distance-determinate but that they are determinate *in the wrong way.*

At first this seems like a reasonable view. But really it is very arbitrary. As Tyler Burge[429] said, a perception to the effect that P is necessarily *prima facie* evidence that P. If you *see* a red chair in front of you, that is necessarily prima facie evidence that there is a red chair in front of you. In the story we are telling, you wake up every morning with goggles of a different level of magnification attached to your eyes, and you never wake knowing what the

level of magnification is. Under those circumstances, your initial sensations are not even *prima facie* evidence of facts relating to the sizes or distances of objects. Given only your initial sensations, you would be equally justified in believing that you were seeing 100 ft object 100 ft away or a ten thousand foot object 10,000 feet away. So your initial sensations wouldn't be *prima facie* evidence of either state of affairs – or, by analogous reasoning, of any other state of affairs relating to the distance or size of the object in question. Thus your initial perceptions wouldn't *tell* you that the object was this as opposed to that size or distance; you wouldn't, veridically or non-veridically, *see* the object as being of this or that size or distance.

The objector's point can be confuted in another way. Again, suppose you wake up wearing goggles that strengthen your vision four-fold, but when you wake up you don't *know* what magnification these goggles have. After you 'acclimatize' to your goggles, you will simply *see* objects as having the dimensions that they in fact. (We are leaving aside cases of objects that are extremely distance or extremely small. One doesn't really *see* a star to a million light years away; one doesn't *see* it, veridically or non-veridically, to be *any* determinate distance away from one. And of course one doesn't *see* a proton, veridically or non-veridically, to have this or that size; for one doesn't see protons at all.) After this acclimatization, when you look at an object that is ten ft tall and twenty ft away, you won't *first* see it as forty ft tall and eighty ft away; you will see it for the size and distance that it actually has – just as the veteran of the tie-shop, even while in the tie-shop, will *see* ties for the colors they actually have: he won't *see* a burgundy tie as being orange and *then* consciously figure out, on the basis of his background knowledge, that what he is seeing is *in fact* burgundy. He will *see* the tie for the color it has.

Let us refer to the person who has acclimatized to his goggles (with

magnification × 4) as Post-Bob, and let us refer to that same person, after waking up but prior to that acclimatization, as pre-Bob. We've agreed that Post-Bob *sees* objects as having the sizes and distances that they actually have. (That is, Post-Bob has size-right and distance-right perceptions *in so far as any one does*. Whether we're dealing with Post-Bob or with a 'normal' person of 20/20 vision, there will inevitably be perceptual vagueness and error, especially where extreme sizes or distances are concerned). Now the objector says that pre-Bob simply *sees* objects as being 4 times their actual size: so his perceptions are not size-indeterminate, but systematically size-*wrong*. But that puts the objector in an awkward position. Pre-Bob's visual *sensations* don't change as he metamorphoses into Post-Bob. But it is very hard to see how a determinately *false* perception could become a determinately *correct* perception simply as a result of a change in background knowledge. It seems that if, when you woke up, you had perceptions that were systematically *false*, as opposed to indeterminate, a change in background knowledge wouldn't result in *veridical* perception: it would simply enable you to intellectually[430] *correct* for your *non-veridical* perceptions. But this isn't what happens: Post-Bob *sees* things the right way; he doesn't *see* things the wrong way and then make a mental note to himself along the lines of 'I see that tree as being 80 ft tall, so it must really be 20 ft tall'. No – he *sees* the tree as being the height that it *actually* is.

So it is hard to see how systematically *false* perceptions could be turned into veridical ones through a change in background knowledge. But it is easy to see how *indeterminate* perceptions can be turned into determinately *veridical* ones once background knowledge is filled in. If one's perceptions are indeterminate (in respect of, say, size or distance), then there is a kind of openness to correction built into them: there is a parameter in it that has yet to be set and that consequently *can* be set. But if a perception is determinately

false, it is frozen; background knowledge won't turn it into a veridical one; the perception will continue to give one its wrong message (though, armed with the right background knowledge, one will be able to intellectually superimpose the right message). So it is very hard to see how a determinately *wrong* (or, to use Quine' word, a *falsidical*) perceptions can become veridical through a change in background knowledge. But it is very easy to see how a change in background knowledge can turn indeterminate perceptions (for example, size- or distance-indeterminate perceptions) into determinate ones. We are taking the latter position. The objector is committed to the former. Therefore the objector is wrong. Pre-Bob's perceptions are size- and distance-*in*determinate, not size- and distance-*false.*

Let us sum up thus far. Under non-standard conditions of perception, one's sensations do not by themselves apprise one of the sizes or, therefore, the distances of objects. What does this is those sensations operating *in conjunction* with a great deal of background knowledge. Absent this back-ground knowledge, one's visual sensations are simply size- and distance-indeterminate.

Exactly similar reasoning shows that, under non-standard conditions, our perceptions are mass-indeterminate, temperature-indeterminate, velocity-indeterminate, and so on. Given *any* perceivable property P, under non-standard conditions, our sensations do not by themselves tell us to what degree, if any, x has P; only our sensations operating in conjunction with background knowledge (about the conditions of perception) do that. Under non-standard conditions, our sensations, taken by themselves, are temperature-indeterminate, mass-indeterminate, and so on. They become temperature- and mass-determinate *only* in so far as they are put in the context of a great deal of back-ground knowledge. In general, under non-standard conditions, perceptions are indeterminate. For any kind of

perceptible property P – P could be size, distance, color…-- it would be possible to create non-standard perceptual conditions such that, under those conditions, one's perceptions were P-indeterminate. And, under such conditions, what makes one's perceptions become P-*de*terminate is the acquisition of *background* knowledge – knowledge of facts about the conditions of perception.

Another way of putting this is to say: for any kind of perceptible property P, it would be possible to create circumstances in which one's perceptions are P-determinate *only with the help of background knowledge.* Under such circumstances, when one saw an object O as having P (or as having P to degree n), a great deal of background knowledge would be implicated in that perception. Consequently one's perception of O as having P would not be *foundational*; it would be built on top of, so to speak, a great deal of background knowledge about the conditions of perceptions. Under those circumstances, it would not be *given* to one that O was P; not even a *false* message to the effect that O was P would be given to one. One would be 'given' (if that is the right word) only a non-epistemic sensation; one's background knowledge would be needed to assign that sensation a P-value.

So far we've spoken about *non*-standard perceptual conditions. But it is easy to show that what we've said applies equally to standard perceptual conditions. For standard conditions are just non-standard conditions that one has got used to; and the distinction between standard and non-standard conditions is quite arbitrary. In fact, *no* conditions are quite standard; if we make conditions *perfectly* standard – the way they might be in a surgical theatre or in the kind of room where they make computer chips – the result is a situation where the perceptual conditions are so good that they are not standard any more. And one is *always* doing in a small way what, in the anecdote above, Bob was doing in a big way: one is always making

allowances for background knowledge.

Let us once again think about Bob. One morning Bob wakes up fitted with goggles of magnification ×2. But this time the goggles stay on for years and years. This time, since Bob wouldn't be able to take off his goggles, it would not be an option for him to compare his goggle-produced perceptions with his non-goggle perceptions: so he couldn't' he would no longer have that way of figuring out *what* his visual modality was telling him. But he would still acclimatize. From his pre-goggle days, he obviously knows the sizes of objects – of stop signs and baseball bats, and so forth – and he could use these as benchmarks. Initially, he would *consciously* correct for his perceptions by using these benchmarks; he see an object, not know how big it was, compare it to a baseball bat, see that it was the same size, and *then* know how big it was. But eventually, no doubt, Bob would simply *see* objects for the sizes they were. (At least in so far as anybody does: nobody's vision is without significant indeterminacy in many respects; one doesn't *see* very distant or very large objects as being specifically this or that size or, derivatively, *as* moving with this or that speed.)

Of course, in some ways, Bob's vision will be much better than that of an ordinary person's, and in other ways, it will be worse. (He will have superb far-sight, but terrible near-sight.) But this doesn't matter to our discussion. What matters is this. Bob's ×2 vision will *become* standard for Bob – just as your vision, with its ×1 magnification, is standard for you. It is clear, given what we've so said so far, that Bob's perceptions acquire determinate content (at least in respect of distance and size and related properties, like speed) only with the co-operation of background knowledge. Not only does Bob's ability to acquire *knowledge* (of distance, size, and related properties) presuppose a great deal of background knowledge. So during this ×2 phase of Bob's vision, sizes and distances (and related properties) would not be *given* to Bob;

his perceptions would be size- and distance determinate only by courtesy of a great deal of background knowledge. So distances and sizes would be given to Bob in the sense that he could, within limits, see what sizes and distances objects had; they would be *made known* to him. But sizes and distances would not be given in the sense that his distance- and size-perceptions would be *foundational*; he would have size- and distance-perceptions *only* because there was a lot background knowledge at work. Of course, in so far as a hefty body of background knowledge must lend its help to make perceptions represent things as having determinate sizes and distances, to that extent sizes and distances are not *given* to one; one has to *earn* them, so to speak. So in a special, but by no means arbitrary, sense of 'given', distances of and sizes are *given* to Bob during this ×2 phase of his visual career.

But now suppose that, after all these years, Bob's revision reverts back to how it used to be. So it is no longer ×2, it is ×1. His vision now would *not* be standard for him, even though it used to be. He would have to do now what he had to do before when he first shifted to ×2 vision. He would acquire *new* background information, and this new background information would be implicated in making his *new* perceptions be size- and distance-determinate. So, just as distances and sizes were not *given* to Bob in his ×2 phase, so they are not given to Bob in his *new* ×1 phase. Exactly similar considerations apply.

But what about Bob's *old* ×1 phase? Maybe, it might be thought, sizes and distances were just *given* to him then. But it would be highly arbitrary to say that. There are two reasons for this. First, it is very hard to believe that Bob's *new* ×1 perceptions function in a completely different way from his *old* ×1 perceptions. We know that his new ×1 perceptions are size- and distance-determinate only by courtesy of a great deal of background knowledge concerning the conditions of perception. But if, during Bob's *old* ×1 phase,

distances and sizes were just *given* in perception (in our special sense of 'given'), then the epistemic liason between Bob and the world constituted by Bob's *old* ×1 perceptions would very different from that constituted by his *new* ×1 perceptions. But there seems no reason to believe this.

There is a second reason to think it arbitrary to say that during Bob's *old* ×1 phase distances and sizes were *given* to him (in our special sense of 'given'), whereas during his ×2 phase they are not given. We've been discussing extreme changes in the conditions of perception. But one's visual career is ordinarily characterized by much smaller, but still very genuine, changes perceptual conditions. And what we said in connection with artificially dramatic change in these conditions is just as true of the smaller versions of such changes that actually occur. One can scarcely walk ten feet, or even rotate one's head, without the lighting conditions changing. But usually (though not always) one doesn't have to make a conscious adjustment. One automatically makes the adjustments, and sees green objects *as* green and red objects *as* red.

The kind of sensation that mediates a perception of a particular shade of red is heavily dependent on highly variable conditions. One doesn't realize this until one tries one's hand at painting, and has to reckon with the setting of the sun or the uneven light cast by a lamp. To reproduce, in painting, two different objects that one *sees* to be of a single shade of red, one may have to use two very different shades of *paint.* When one looks at the paint*ing*, one sees the two objects as being of the same shade of red – just as the artist, when he was having the actual perception which he reproduced in the painting, saw the two objects as being the same shade. But when one looks at the *paint* – not the paint*ing*, not the things *depicted* by means of the paint – one sees two *different* shades of red. The differences in *shade* correspond to differences in *quale*, in *sensation.* The identity of color of the *objects*

represented in the painting corresponds to the identity in representational content of the *perceptions* reproduced in the painting.

Thus one makes innumerable adjustments to ever changing perceptual conditions. So one is perpetually adjusting to non-standard conditions. We can only assume, by arguments analogous to those given earlier, that one's background knowledge is implicated in these perceptions. Colors (for example) are not *given* to one; what is given to one is a non-epistemic sensation – a sensation with no determinate color- (or distance- or size-…) content. What *does* constitute a color- (or size-…) determinate perception is that sensation *in conjunction with* a great deal of back-ground knowledge.

Or course, arguments exactly analogous to those that we've given in connection with size, distance, and color apply to mass, temperature, and any other kind of perceptible properties: perceptible properties are not *given* (at least not in our special sense of the word). Whenever one has a perception that is mass- or size- or temperature or color-determinate – that is, whenever one has a perception that even *purports* to tell one what color or temperature or mass… an object has – that is, in part, because one's quale *sensations* are interacting with a great deal of background knowledge. The sensations by themselves are not color- or size- or temperature-determinate. The sensations themselves – the 'raw feels' or quale -- are not representational. The sensations *plus* the background knowledge is what is representational.

Of course, a perception is, by definition, representational; a perception necessarily is either size- or color-… or mass-determinate; a perception necessarily gives one *some* kind of message and is thus not *wholly* indeterminate. So what we have seen is not, obviously, that *perceptions* are not representational (that would be absurd) or that facts about the world are not *given* through *perceptions* (that too would absurd: there is no other way that facts about the external world are given to us). What we have seen is that

the very existence of any single perception *presupposes* the co-operation of background empirical knowledge. (This raises the question how we can ever have a *first* perception: a question handles very delicately. There isn't room to address this problem here. But I don't think it constitutes any significant obstacle to the anti-foundationalism sketched here.)

The 'myth of the given' is that a given sensation *by itself* constitutes some kind of awareness of an external reality, and that no background knowledge is presupposed. Of course, if this were the case, then sensations would be epistemic rock-bottom: they would constitute the foundations of all knowledge. In sensation, certain bits of information would be *given* to us – in our special sense of 'given' – and all knowledge would be built on top of this given information. (And, in this context, to say that piece of information is 'given' to us in a sensation is to say that no *background* knowledge would be implicated in the acquisition of this information.)

Actually, what we've seen is not only that no *knowledge* is 'given' (in our special sense), but that nothing *representational* is given: background knowledge is implicated in the generation of anything that even *purports* to give information about the world. So not only is no *knowledge* given; *nothing* even representational is given. There are thus two myths of the given: one is that *knowledge* is given; the other is that there are perceptions *in a vacuum* of background knowledge – in other words, that in some cases one's sensual experiences do not need the co-operation of background knowledge to constitute perceptions (or even misperceptions). I submit that intentionalism is guilty of the second 'myth of the given' and thus *a fortiori* is guilty of the first.

We've seen that the perception results when the quale interacts with background knowledge: before that happens, there is nothing representational – the quale is size- and color- and temperature--…indeterminate. So there is

an extremely *strong* sense in which nothing is given: nothing that even *purports* to give true information about the world exists *without* the co-operation of background knowledge. *A fortiori* there is no *actual* awareness of what is going in the external world without the co-operation of background knowledge. Sensations by themselves are not just epistemically, but *representationally*, inert; they are, as we put it, 'indeterminate' until background knowledge sets the relevant parameters. And different bodies of background knowledge can 'set' these parameters differently.

All of this is very inconsistent with intentionalism. Intentionalism says that there cannot possibly be a difference in phenomenology unless there is a difference in representational content. Intentionalism thus assimilates the phenomenological to the representational. Given that there cannot *possibly* (in any possible world) be a difference in phenomenology without one in representational content, it follows phenomenology is realized or constituted by something have representational content. (Otherwise the supervenience of the phenomenology would be inexplicable.) And it also follows that, if *any* bit of phenomenology was non-representational, it would have to be the *only* kind of phenomenal content was non-representational. (For if there were *two* non-representational types of phenomenal content, then there would be a difference in phenomenal content that was, vacuously, *not* matched by a difference in representational content: obviously two bits of phenomenal content that represent *noting* don't represent *different* things.)

But the picture we've sketched contradicts these contentions in a number of way. First, we've seen that phenomenology is *not* typically realized by (let alone identical with) something representational. Perception involves the co-operation of sensations (quale) *and* back-ground knowledge; in such cases, the sensation is not inherently representational, but becomes so only with the help of background knowledge. Thus at least some sensations are not *by*

themselves representational. So this contradicts the contention that phenomenal content is always realized by something having representational content.

Also, it would seem that the inherently non-representational qualia that are involved in one perception can be dramatically different from those involved in another perception. So this contradicts the idea that there can be *at most* one kind of non-representational phenomenal content. Thus intentionalism is contradicted on two levels.

To sum up, the intentionalist maintains that a quale *by itself* typically constitutes a representation; for the intentionalist maintains that the thing having phenomenal content is identical with, or realized by, the thing having representational content. (Thus, for the intentionalist, the thing having phenomenal content cannot possibly fail to be representational– just as the statue cannot fail to be, say, wet if the clay that realizes the statue is wet.) But to say this is fall squarely into the myth of the given: the spurious belief that perceptual representations *do not* implicate background knowledge and sensations *by themselves* are epistemic.

In one place, Sellars himself writes: "Sensations are no more epistemic in character than are trees or tables…"[431] My guess (though the context of that quotation makes it unclear whether this guess is correct or not) is that he had in mind points like the ones we've just stated.

We arrived at these points by doing a great deal of theory. But they are borne out by examples. Under certain lighting conditions, you see a tie – literally *see* it – as being burgundy. Under different lighting conditions, in having a phenomenologically identical perception, you see – literally *see* that tie as maroon (or, in any case, some color other than burgundy). You see a tree as being moderately tall and being (about) twenty feet away. But, given different background knowledge, a phenomenologically identical

perception could tell you that you were seeing a gigantic tree that was (about) a hundred feet away.

Suppose that Bob is in his × 1 phase, and that P is a bit of phenomenal content that mediates a perception of a small tree nearby. When Bob is in his ×2 phase, or his ×10 phase, a bit of phenomenal content qualitatively just like P will mediate a perception of a large but distant tree. The *same* bit of phenomenal content was associated with different representational contents. This is inconsistent with intentionalism.

Intentionalists tend to defend their position by seizing on facts like the following. Suppose that in circumstance C1, some sensation is involved in seeing that a tie is burgundy. And suppose that in C2, some qualitatively *different* sensation is involved in seeing that a tie is burgundy. (This will happen if the lights are dimmed somewhat, or if the over head light is replaced with a less pure, more yellow light: it could happen under any number of circumstances.) Of course, *prima facie* this seems to favor the anti-intentionalist: different representational contents, same phenomenal content. But the intentionalist always responds by saying:

The two perceptions *don't* have the same representational contents. In one case, part of what you are seeing is a light casting an impure yellowish light. In the other case, you may not be seeing that, and you may be seeing a light that is casting a very pure, white light. Or in the one case, the tie may be better *illuminated* than in the other case. So the two qualitatively identical perceptions are representing *different* things: in the one case a *well* illuminated patch of burgundy is represented; in the other, a poorly illuminated patch of burgundy is represented. And *that* is why there is a difference in phenomenal content: for *there is* a difference in representational content.[432]

But this doesn't apply to Bob. Depending on whether he is in his ×1 phase or his ×10 phase, a given a bit of phenomenal content may mediate perceptions with very different representational contents. So there is *sameness* of phenomenal content and a *difference* in representational content. The same bit of phenomenal content is deployed once in seeing a small nearby tree, and another time in seeing distant large tree.

These considerations help us identify the shortcoming in a well known argument for intentionalism due to (or at least stated by) Alex Byrne. Byrne gives the following argument for intentionalism.

Suppose you are looking at a blue chip. Your seeing the blue of the chip obviously involves your having some kind of phenomenal content. Now imagine a change in the phenomenal content that is involved in your seeing the blue of the chip. If there is such a change, then there is necessarily a change in the color that the chip is represented to as having; it goes from being represented as blue to being represented as (say) green. The same point (mutatis mutandis) applies to *any* perception. If you imagine a change in the phenomenal character of your gustatory perception of the honey, then the honey goes from being represented as sweet to being represented as (say) bitter. One last example: suppose you hear a certain pitch *as* a middle C (you have perfect pitch); if the phenomenal content involved in your hearing that pitch is changed, then *ipso facto* you no longer hear that sound as a middle C but as (say) a middle D. So: no difference in phenomenal content without a different in representational content.[433]

But there are obvious counter-examples to this. You can feel the roundness of the ball and also see it (a bat could hear it, a mole could smell it). But these perceptions have *no* phenomenal overlap, despite considerable (though my no means total) representational overlap. You can *see* distances, and you can also learn distances kinesthetically. Obviously a visual perception has no phenomenal overlap with a series of kinesthetic sensations, despite a certain overlap in representational content.

Also non-occurrent memories have *no* phenomenal overlap with perceptions whose representations they store (or can store, in principle) in their entirety. Total overlap of representational content: no phenomenal overlap. (Of course intentionalism is defined as the position that two *experiences* cannot differ in phenomenal content without differing in representational content. So it may see that intentionalism doesn't say anything about whether a non-occurrent memory can have the same representational content as, say, a perception without having any phenomenal overlap with it. But this is not the case. Consider a perception P with phenomenal content C and representational content R. Intentionalism says that having P has C wholly *in virtue* of having R. So, for intentionalism, having R is a *sufficient* condition for having P. So, if intentionalism, then nothing *other* than having R is needed to have C. So intentionalism is committed to the view that anything has P as long as it has R. But a non-occurrent memory has R but not P.)

Byrne's argument proves *at most* that *within a single sensory modality* there cannot be phenomenal variation without representational variation. But this thesis is false, it turns out, if we consider different sensory modalities or memories.

But we've seen that Byrne's thesis is false *even* within a single modality. At

the same time, it is not an accidental fact, surely, that within a single modality, differences in phenomenal content *tend* to co-vary with differences in representational content. And, admittedly, the most obvious way to explain this is by advocating some kind of intentionalism. So we owe Byrne an non-intentionalist explanation of this co-variation.

What Byrne has shown is not that phenomenal content co-varies *tout court* with representational content, but that phenomenal content co-varies with representational content *all other things being equal*. And I submit that, in this case, the 'other things' consist of one's background knowledge.

Obviously phenomenal content is involved, in an essential way, in our perceptions; and, obviously, perceptions are involved in an essential way, in our acquiring knowledge of the external world. Now *if* phenomenal contents are to be of assistance in our acquiring empirical knowledge, then presumably they must be paired off in a *consistent* with way their (proximal) causes. If a given kind of sensory disturbance (i.e. a given kind of disturbance of one's retina or epidermis or eardrum…) result now in one kind of phenomenal content, and now in some totally different kind of phenomenal content, then one's phenomenal contents wouldn't be much use in helping us acquire knowledge of the external world. We would be lost. (We can tolerate *occasional* shifts of this kind. But these shifts cannot be the norm: they must occur against a background of stability.) So there must be a reasonably consistent relationship between our phenomenal contents and their (proximal) causes. There must be some rule R that pairs off tissue disturbances (of our sensory organs) with phenomenal contents (and does so in some not ridiculously Byzantine way).

But is there only *one* possible value of R? Do your perceptions have to be associated with their causes in *exactly* the way they are *in fact* associated with them to tell us what they do about the world? Under certain conditions,

sound-wave oscillations of 440 Hz cause a certain kind of phenomenal content (call it P1); under those same conditions, sound-wave oscillations of 480 Hz cause a certain kind of phenomenal content (call it P2), and so on. Let R1 be the rule that pairs off phenomenal contents with disturbances of your eardrum in that particular way. Must *R1* be the operative rule of projection? Would a being whose auditory system involved R2 (some rule other than R1) necessarily be mis-hearing everything? It seems that, if R2 is as *consistent* as R1, it can be as good a rule of projection, i.e. it can be involved in the generation of perceptions that are as accurate (if not more so) than those had by a being with R1. In fact, R2 might be better than R1 in this respect. R1 may assign the same phenomenal content to sound-wave oscillations of 440 Hz that it assigns to those of 445 Hz. (Human beings are not even *remotely* capable of registering such minute differences, though other creatures are. So R1 might the rule operative in human hearing.) If R2 is *more* sensitive to differences in sound-wave oscillations – and clearly non-human creatures are better able to hear differences in pitch than we are (and this presumably involves having different *phenomenal* contents in response to different pitches in cases where we would have the *same* phenomenal content in response to both pitches) – and R2 is just as consistent, then it seems extraordinarily arbitrary to say that creatures whose hearing-system uses R2 are just mis-hearing all the time. But that is just what the intentionalist is committed to. Obviously humans hear things the right way, at least some of the time. And obviously our hearing things involves having certain phenomenal contents in response to sound-wave oscillations. The intentionalist says that a creature that has a *different* phenomenal response to the same sound-wave oscillations. The intentionalist *also* says that a difference in phenomenology means a difference in representational content: so those other creatures are *hearing things as being different* and thus – since

we are hearing things the right way and they are hearing things differently – as hearing them *wrong*. To clarify this: Suppose a *human's* hearing (under certain circumstances) a middle C involves having phenomenal content P, and that human a C correctly under those circumstances. *For an intentionalist*, a creature that has phenomenal content P* (different from P) under those same circumstances is *not* hearing a middle C; it is hearing some other pitch. So that creature is hearing the wrong thing. Thus if that human's audition involves R1, then every other right-hearing creature's hearing must involve R1 as well. No other correspondence rule, no matter how fine-grained (even if it is much more nuanced than ours) or consistent will do.

But again this seems deeply arbitrary. If another creature's phenomenal responses to sound-wave oscillations are just as consistent as ours and just as fine-grained, then by any reasonable conception of information, it can acquire as much information about the world through hearing as we can; and it then seems arbitrary to say that it systematically mis-hears just because it is employing R2 and not R1. And if a creature's hearing involves a *more* fine-grained, and more internally consistent rule than ours, then by any reasonable conception of information, it can acquire *more* information (through hearing) than we can about the world; and it then becomes all the more arbitrary to say that such a creatures is always hearing things the wrong way just because it uses R3 not R1. And there is no doubt that there are, in principle, rules *other* than the one operative in human hearing that *are* more fine-grained than ours and as consistent than ours. (to say that a rule R is more 'fine grained' than a rule R* is to say that, for some n, if sound-waves differ from each by less than n, but still differ, R will, whereas R* will not, pair off those sounds with different phenomenal contents. There can be no doubt that, by this definition, there are creatures whose hearing involves much more fine-grained rules of correspondence than ours – bats, dogs, rabbits.) So there seems no reason in

principle to *deny* that creatures that hear *better* than us might have different phenomenal contents from us in response to a given sound-wave under a given circumstance. But intentionalism says *no difference in phenomenal content without a difference in representational content.* And the motivation for this view is that the phenomenal reduces, after a fashion, to the representational. But apparently it does not. [434]

I would suggest that what Byrne has shown is that *as long as* a given rule of projection is operative, then the phenomenal 'tracks' the representational, i.e. if perceptions have different phenomenal content, then they have different representational contents. This is to be expected since an *inconsistent* rule of projection – one that assigned *different* phenomenal contents to the *same* sound (heard under the same circumstances) – would not be of much use (in learning about actual sound-phenomena). But what Byrne doesn't register is that *if* the rule of projection is changed, then the representational value of a given phenomenal content can change. Suppose that Bob is a human being who has (by human standards) good hearing. Let R1 be the operative rule projection. Suppose that neurosurgeons change Bob's auditory system. Now the rule of projection that is operative in his hearing is the one operative in a bat's. Obviously Bob's hearing a lot better than it used to be. But given what we said a moment ago, it seems highly likely that, in this new and improved version of Bob, a phenomenal content that *used* to be involved in hearing (say) a middle C is *now* involved in hearing a middle D. So when the *rule of projection* is changed, Byrne's thesis no longer applies: phenomenology tracks differences in representational content only *given* a certain rule of projection. In any case, this seems plausible.

We've said a lot about the role of 'rules of projection' in the assigning of representational contents to phenomenal contents. What we've said on this

matter is consistent with Sellars' attack on the myth of the given. Sellars' attack, we've seen, comes in two strengths. Weak version: it is not typically true that qualia, by themselves, constitute *awarenesses* (knowledge) of anything external – something extra is needed to make a quale be *epistemic.* Strong version: it is not typically true that qualia, by themselves, constitute *representations* – something extra is needed to give them representational content.

We've seen reason to hold that both versions are correct. But we seem to have a problem. We've apparently given two different answers to the question 'what is that extra something which makes a non-representational (or non-epistemic) quale be representational?' One answer we gave is: *background knowledge.* The other answer is: *a rule of projection.*

To deal with this, we need only be clear on what we've said so far. The rule of projection doesn't assign *representational contents* to disturbances of our sensory surfaces. The rule of projection assigns only *phenomenal contents* – qualia – to such disturbances. It is *background knowledge* that assigns representational contents those qualia. So the 'rule of projections' that we've been talking about have an altogether different function from background knowledge, though of course the two are part of a co-operative enterprise.

Conclusion

Intentionalism says that, typically, qualia *by themselves* have determinate representational content: background knowledge isn't needed to *endow* qualia with representational content. And, of course, in so far as qualia have representational contents *ab initio*, background knowledge cannot affect what representational content they have. For *if* background knowledge *did* have a

hand in endowing a quale with representational content, it could only be by way of some openness, some representational indeterminacy, on the part of the quale; and, by definition, a quale with a 'fixed' representational content has no such indeterminacy or openness. Thus the picture given to us by the intentionalist is that, typically, qualia don't *need* background knowledge to fix what they represent, and they won't *let* background knowledge fix what they represent. Qualia thus become *foundational*, performing their representational and epistemic roles in splendid isolation of any epistemic context. But it is just this kind of foundationalism that Sellars (and, after him, many others) have shown not to be tenable. Intentionalism is a strong form of foundationalism; it is guilty of the myth of the given, in both its forms. Therefore *either* intentionalism is to be rejected *or* the anti-foundationalism of Sellars and others – now so widely held – must be false.

Burge, Tyler. 1993. "Content Preservation". *The Philosophical Review*, Vol. 102, No. 4, (Oct., 1993) 457-488.

Byrne, Alex . 2001. 'Intentionalism Defended'. *The Philosophical Review*. Vol. 110, No. 2 . (April 2001).

Sellars, Wilfred. 1963a. "Empiricism and the Philosophy of Mind". In *Science, Perception and Reality*. New York: Routledge and Kegan Paul.
-- 1963b. "Some Reflections on Language Games". In *Science, Perception and Reality*. New York: Routledge and Kegan Paul.

Uniquely Individuating Descriptions

Let me start by stating the picture I wish to call into question

Many a person can refer to Socrates – can use the word "Socrates" or some synonym with meaning and understanding – but not be able to produce a description that applies to him and not also to many others. One can use the word "Socrates" with understanding and *not* know the truth of any proposition of the form: *somebody uniquely has phi and "Socrates" refers to x.* When asked who "Socrates" refers to, they may say "he was a great philosopher of ancient times" or "he wrote the Nichomachean Ethics"; and that may exhaust what they have to say by way of an answer to your question. What they "know" may be excessively general ("he was a great philosopher of ancient times") or it may simply be false ("he wrote the Nichomachean Ethics"). In both cases, they will lack knowledge of any truth claim of the form: *somebody uniquely has phi and "Socrates" refers to phi.* They will lack knowledge of the right kind of *uniquely individuating description.*

But such people may still be able to produce sentences of the form "… Socrates…", and really *mean* them. So apparently one can refer to Socrates, and think about him, without knowing of any description that singles him out.

The purpose of this essay is to argue for two parallel points. First, in order to refer to an object, one *does* have to know of some description that singles it out: one has to have knowledge of the right kind of *uniquely individuating description*. Second, in order to *think* about an object, one *does* have to know of some description that singles it out. Without knowledge of a *uniquely individuating description*, there is no reference and there is no conception. In many cases, the ability to refer to, or think about, an object *does* involve one's being causally connected to the object. But the causal connection,

though essential, nonetheless has a subordinate role: it facilitates conception and reference *by way of* being an indispensable part of one's knowledge of a uniquely individuating description.

I have no doubt that Kripke's major point about proper names is quite correct. "Socrates" is not a disguised definite description. Further, I concede that *part* of what makes tokens of "Socrates" refer to Socrates is (or, more accurately, *can* be) that those tokens stand in a certain causal relation to Socrates himself. Finally, I concede that part of what enables one to think about Socrates is (or, more accurately, *can* be) that one stands in a certain causal relation to the man. But I will argue that these causal relations are always embedded *in* knowledge of a description that applies to Socrates and Socrates alone – are embedded in knowledge of a *uniquely identifying description.*

None of Kripke's important semantic insights require an abandonment of the position that, if one is to think about x, one must know of some description or concept that applies uniquely to x. Indeed, when we delineate what Kripke says, we will find that, on the contrary, Kripke's story *requires* that one always have such knowledge.

I should make it clear what I will and will not be attacking. I believe that direct reference theory is correct – that the sole semantic value of "Socrates" is Socrates. But semantics is not epistemology. Given only that "Socrates" is a label, and not a description, it does not follow that I can *think* about Socrates without knowing a description that singles him out. And, I will argue, thinking about Socrates *does* require knowledge of such a description. Indeed, being able to *refer* to Socrates requires knowledge of such a description – notwithstanding that "Socrates" itself has no descriptive content at all.

A consequence of these findings is that a popular and seemingly plausible

analysis of conception is quite false. In fact, much contemporary philosophy of mind and epistemology is a projection of views that *seem* to have support in *Naming and Necessity*, but in fact have none.

Let us start with a very quick summary of some of the relevant points in *Naming and Necessity.* According to Russell, "Socrates" is synonymous with a definite description; this might be "the great philosopher of antiquity who died of hemlock poisoning". To *understand* the word "Socrates", and thus to understand sentences of the form "…Socrates…", one must obviously know the semantics of "Socrates". So one must know that "Socrates" is synonymous with that definite description. Consequently, anyone who uses the word "Socrates" with understanding is able to produce a description that applies to Socrates and Socrates alone. In general, if one knows the semantics of a referring term, then one knows some description that applies uniquely to the referent of that term. If I can use the word "Gilgamesh" with understanding, that means I know of some description that singles out that individual. If I can produce no such description, then I cannot really know what is meant by sentences like "Gilgamesh was wise".

Supposedly, Kripke demolished this view. Consider *any* description of Socrates that applies to him uniquely, i.e. any description that truly singles him out. Here is an example: "the greatest philosopher to have died of hemlock-poisoning". One can understand sentences of the form "… Socrates…" without knowing that the referent of "Socrates" consumed hemlock. One can understand "Socrates was wise" without knowing that "Socrates" picks out somebody who uniquely satisfies the description: *greatest philosopher to have died of hemlock-poisoning.* This point holds *mutatis mutandis* for any description that singles out Socrates. Consider the description "the person who figures as the protagonist in most of the platonic dialogues". One can understand the sentence "Socrates was wise" without

having the slightest idea that he was a protagonist in any philosophical dialogues. Given *any* description D that applies uniquely to Socrates, one can understand "…Socrates…" without knowing that D applies to the referent of "Socrates".

This point has many apparent consequences. Some of these are semantic in nature. Others are epistemological.

Here is one of the semantic consequences. "Socrates" is *not* synonymous with any definite description. Given *any* description D, one can know that "Socrates was wise" is true without knowing that somebody uniquely instantiated D and that any such person was wise. So it isn't tenable to maintain that "Socrates" is synonymous with "the greatest philosopher to have died of hemlock-poisoning" or with any other definite description.

Here is one of the *epistemological* consequences. One can *think* about Socrates without knowing of any description that singles him out – without knowing a UID.

There are other apparent consequences, both for semantics and epistemology. Let us start with a question: Why does "Socrates" refer to Socrates and not, say, Plato? Russell's answer was this: "Socrates" is synonymous with a description that is true of Socrates, and Socrates alone. Kripke demolished this answer. So how does *Kripke* answer our question? His answer is basically this: "Socrates" refers to Socrates because a certain kind of *causal* relation holds between tokens of "Socrates" and Socrates himself (or, more exactly, states of affairs involving Socrates).

Kripke has more to say on this matter. Somebody – say Smith -- *saw* Socrates. So Socrates' presence *caused* Smith to have certain visual experiences. Smith then gave a name to Socrates. Smith was able to pass that name along to people who did not themselves see, or otherwise sense-perceive, Socrates. Of course, for Smith to pass along that name to someone,

that person had to be causally connected to Smith. The process just described is repeated over and over. I myself am the 521st installment in a sequence of events like the one described. Inevitably, Smith's original term for Socrates has undergone phonetic distortion. What I have been left with is the sound-type "Socrates". When I produce instances of that sound-type, my words refer back to Socrates precisely because I am connected to Socrates by way of a chain like the one just described. When I say "Socrates", my words refer to Socrates because they are causally connected to him in a certain way. In general, if an expression E refers to an object O, that is because tokens of E stand in a certain causal relation to states of affairs involving O.

How do you learn what a given word refers to? Ultimately, there are two ways: by description and by ostension.

You are young; so you have never heard the name "Socrates". Somebody says to you:

(*) " 'Socrates' is the name of the most famous philosopher to die of hemlock poisoning."

Here the word "Socrates" has been defined for you in a *descriptive* manner. In effect, you were told that somebody O uniquely satisfies a certain description -- *most famous philosopher to die of hemlock poisoning* – and that "Socrates" refers to, or names, O.

On the face of it, there seems to be a second, entirely distinct way of learning what a term refers to: you can be given an *ostensive* definition.

You are in a crowded market place. You see a man haranguing a crowd. Your companion says:

(**) “*that* person over there [pointing to the orator] is named ‘Socrates’”.

My own view is that ostensive definition actually collapses into a kind of descriptive definition. I will argue for this later on. But, of course, this is not a point that can be taken for granted. Right now, we need only register the fact that one can learn what a term refers to through descriptive and also through ostensive definition.

When you give a descriptive definition of a referring term, the descriptive information must be taken to have *wide-scope.* Otherwise the definition turns out false.

(*) defines “Socrates” by means of a description: *most famous philosopher of antiquity to die of hemlock poisoning*. But the descriptive information in (*) is used to *fix the referent* of “Socrates”; that descriptive information is not itself part of the *meaning* of “Socrates”.

The following statement is true:

(s1) “Somebody O was a unique most famous philosopher of antiquity to die of hemlock poisoning and, for any predicate *…x…*, *…Socrates…* means:…*O*.…

The following statement is false:

(s2) “…Socrates…” means: *Somebody O was a unique most famous philosopher of antiquity to die of hemlock poisoning and…O…*

"Socrates smoked" does *not* mean: *somebody O was a unique most famous philosopher of antiquity to die of hemlock poisoning and O smoked.*

But there is some object O such that "Socrates smoked" means: *O smoked.*

The descriptive information in (*), and in (s1), is given wide-scope with respect to the "means that" operator. That is why such statements are true. Let us now move on.

It *seems* that, in a certain class of cases, one is able to acquire the ability to use a referring term *without* ever having been given either an ostensive or descriptive definition of it. Suppose you and two other people (Smith and Jones) are talking. All of a sudden Smith and Jones start talking about somebody named "Argo". They never say who "Argo" refers to. They never say: "Argo was the president of our fraternity back in college" or "Argo was my divorce lawyer; and by sheer coincidence he's Smith's racquetball partner." They do not *ostensively* define "Argo"; for he is not present. Smith and Jones simply prattle on about "Argo"; and you are too polite to ask them who that person is.

Nonetheless, as Gareth Evans pointed out, it seems that you *can* still use the word "Argo" in a meaningful way. You can say "it sure sounds like Argo is an awful person"; and you can ask questions like "did Argo mean to do that to you, or was it just an accident?"

Prima face it seems that you can acquire the ability to use "Argo" in a meaningful way *without* ever having been given an ostensive or descriptive definition of it. When you say "it sounds as though Argo is an evil person", you are using the expression "Argo" with understanding; you are not in the same category as a parrot that mindlessly says "Argo was a cad".

Of course, one *could* say this:

Because you haven't been given any kind of definition of "Argo", you

can*not* really use that term with understanding, and you cannot really know what is meant by sentences containing it. When you produce the sounds "Argo sounds like a bad person", you are really making some kind of metalinguistic statement, some statement like: " 'Argo' refers to somebody who is bad". And when you hear sentences like "Argo was a cad", you cannot really understand; what you understand is some related meta-linguistic claim like: "'Argo' refers to somebody who was a cad".

This position cannot be summarily dismissed; and I will not dismiss it. But, for our purposes, it is enough to note that the objector's point *seems* quite false. Cases like the one just described are pervasive. There are many cases where no ostensive or descriptive definition is involved in our appropriation of a proper name. We hear the term being used; and – *presto!* – can use it ourselves. If the objector's point is correct, we are not really *using* any of these terms: we are merely *mentioning* them; when we produce sentences containing them, we are talking not about people, but about symbols denoting people. But this seems false. I have never been given a descriptive or ostensive definition of the term "Charlemagne". But surely I can make statements about Charlemagne, and not just about "Charlemagne".

Let us take stock. There seem to be three ways that a proper name N can be inducted into one's lexicon.

(i) One can be given an ostensive definition of N;
(ii) one can be given a descriptive definition of N; or
(iii) one can just "pick up" the use of N.

Let us consider each category. Suppose that somebody *ostensively* defines "Socrates" for me. I see somebody haranguing a mob. My companion

says: "*that* person is named 'Socrates'".

Under this circumstance, I know that there is some individual x such that x uniquely has certain properties (he is uniquely a person that I am seeing on this particular occasion behaving in a certain way, dressed in certain clothes, and so on), and I know that "Socrates" refers to x.

Ostensive definitions work by way of knowledge of uniquely individuating descriptions. Somebody *points* to some person (or object) O and says: "*that* guy is named 'Fred'". For the definition to work – for you to learn what "Fred" refers to – you must have some way of singling out the person ostended. Suppose your companion points in a certain direction and says "that is Fred", but there are thirty different equally salient objects. In that case, the definition won't work. You must be able to pick out the right object.

Picking out a single object involves knowing a uniquely individuating description. An extreme case will bring out my meaning. Suppose that there are thirty numerically different, but qualitatively different men in a certain place. Your companion Bob goes up to one of them, put his hand on his shoulder, and says "this guy is named 'Fred'". Even though the person just named is qualitatively just like many others, you now of some description that applies to *him* and to no one else. You know that: There is some individual x such that x is uniquely a person whom Bob just picked out – such that x uniquely has the property that Bob just put his hand on his shoulder – and "Fred" names x.

Now imagine a slightly different scenario. You and Bob are in the presence of thirty qualitatively identical individuals. (They are in different places of course.) Bob doesn't actually put his hand on anyone's shoulder. But there is some person x such that Bob *is* able to use an act of pointing to distinguish x from the other people (and objects) in his vicinity and such that Bob tells you that "Fred" refers to x.

Bob's definition works. It works only because you know of some existence claim that Fred, and Fred alone, satisfies. You know that: there is some specific person x such that x uniquely is such that Bob just pointed him out. You also know that "Fred" refers to that person. But your being able to know "Fred" refers to that specific person involves your knowing the truth of some claim of the form: *somebody x uniquely has phi and "Fred" refers to phi.* You know what "Fred" refers to only because you know of some description that singles out Fred: some "uniquely individuating description".

The description is not one that tells you much about what kind of person Fred is. In fact, that description doesn't tell you *anything* about how Fred is psychologically or (leaving aside facts about spatio-temporal location) physically different from other people. The description in question gives you enough information to cognitively single out Fred, and it gives you very little, if anything, beyond that. But that is neither here nor there. In this context, there is only one important point. Your knowing what "Fred" depends essentially on your knowing the truth of some proposition of the form: *somebody x uniquely has phi and "Fred" names x.*

This is not to deny that your knowing what "Fred" refers to involves your having some kind of causal connection to Fred. You *do* have some kind of causal connection to Fred; you *see* him; and this involves your being causally connected to him. You acquire the ability to use the word "Fred" in consequence of (possibly among other things) this causal connection. So your subsequent uses of "Fred" refer to Fred in virtue of, possibly among other things, this causal connection.

But the causal connection is not the primary thing; the causal connection is a component of something more important. It is a component of your knowledge of a uniquely individuating description. How is the meaning of "Fred" given to you? It is given to you ostensively. But, even so, it is, at least

in effect, given to you through an existence claim. You know the truth of some claim of the form: there is some x such that x uniquely has phi and "Fred" refers to x. Now, your knowing the truth of that claim involves your being causally connected to Fred. For the claim in question involved *visual* perception (or, at any rate, some kind of sense-perception); and seeing (or otherwise sense-perceiving) a thing involves being causally connected to it. But *how* is vision (sense-perception) involved? It is involved in exactly this sense: it enables you to know the truth of some existence claim of the kind described a moment ago; sense-perception is involved in the sense it apprises you of the truth of a *uniquely individuating description.* A causal connection to Fred is involved in your knowing what "Fred" refers to: but this causal connection is involved only to the extent that is constitutive of your knowing the right kind of uniquely individuating description.

What we've established so far is this. In *some* cases, not all, we learn what a term refers to through ostension definition. When we acquire the ability to use a term through ostensive definition, our ability to use that term with understanding consists in our having knowledge of some uniquely individuating description of the referent: it involves our knowing the truth of some claim of the form *something x uniquely has phi and N refers to x,* where N is the word in question.

As we saw, there seem to be three ways that a proper name N can be inducted into one's lexicon.

(i) One can be given an ostensive definition;
(ii) one can be given a descriptive definition of N; or
(iii) one can just "pick up" the use of N.

We've already dealt with (i). Let us now deal with (iii). We will soon see why we are skipping to (iii), and not dealing right away with (ii).

Once again, consider the case where you are talking with Smith and Jones. They suddenly starting talking about somebody named "Argo". They are acquainted with Argo, but you are not. At no point in the conversation do they ever say anything like "Argo was my divorce lawyer" or "Argo was the best man at my wedding". They never give you a uniquely individuating description; they never give you a *de*finite description that applies to Argo.

Of course, given that they are producing sentences of the form "… Argo…", it follows that you are, in effect, given many *in*definite descriptions that apply to him. Smith says "Argo is a real cad." So here you are being given an existence claim of the form: somebody x is a cad and "Argo" refers to x. But, of course, many people are cads; so given only that you know the truth of this existence claim, it doesn't follow that you know which *specific* person – which *specific* cad – "Argo" refers to.

The truth is: merely by virtue of the fact that you overheard Smith and Jones using the word "Argo", you *have* been given a *definite* description that applies to Argo, and you *have* learned some true proposition of the form: some thing or some person x uniquely has phi and "Argo" refers to x.

Unlike you, Smith *is* personally acquainted with Argo. Thus, for reasons we've covered, Smith knows some true proposition of the form: some thing (or some person) x uniquely has phi and "Argo" refers to x.

In the course of your conversation with Smith and Jones, you often hear Smith make statements like "Argo was a cad", and "maybe Argo wasn't so bad". On each such occasion, there was *one* person to whom Smith was referring by means of the word "Argo". Suppose that, on occasion *t*, Smith says "Argo was a cad". Obviously the description "x is a cad" is indefinite, and applies to many. But on that occasion, there was exactly one person who

Smith was describing as being a cad; and you know this, merely by virtue of overhearing Smith's statement. So, just by virtue of being colloquy with Smith, you *do* pick up knowledge of some claim like: on occasion *t*, there was exactly one person who Smith described as being a cad, and "Argo" refers to x.

To test the mettle of my thesis, let us make things as difficult as possible for it. Suppose that Smith goes on a rant. He says "Flipper is a cad; Argo is a cad; Ichabod is a cad; Peewee is a cad…" You have never heard *any* of these names before; you have not even heard Smith (or Jones) use any of those names in the context of the conversation in which you are currently engaged. Let us further suppose that you never hear anyone use any of these names again; so on no later occasion are you given any further information about them or their referents. Nonetheless, merely by virtue of overhearing these words on that one occasion, you can appropriate them into your lexicon, and use them meaningfully: you can make statements, and ask question, *about* Argo and Flipper and Peewee; for example, you can say "Peewee and Argo probably didn't have very nurturing parents".

This fact *seems* to confute the theory I just put forth. You hear Smith use the name "Argo" just once; and what Smith is saying about "Argo" is identical with what he is saying about "Peewee" and "Flipper" and all the rest. So, under these circumstances, you do *not* know some true existence claim of the form: *on such and such occasion, there was exactly one person whom Smith was describing as a cad, and "Argo" refers to x.* For that claim isn't true; on the occasion in question, Smith was describing several people as cads.

Despite all of this, I submit, you *still* know of some true existence claim of the form. In fact, you know several such existence claims. First of all, we must take care not to define the term "occasion" too broadly. The term

“occasion” can refer to a three week period or a three nanosecond period. Suppose you go to three-day long party and you meet many people. You can truly say “I met two hundred people on that occasion”. But *within* that occasion, there were other occasions. Perhaps you went to the veranda and met some people there. On *that* occasion you met forty people. And within *that* occasion, there were yet others. There was the time that so and so spilled his drink: referring to *that* incident, you could truly say “on *that* occasion I seriously regretted leaving my house that day.”

Terms like “occasion” and “context” are (pun intended) context-sensitive. An occasion (or context) could comprise nothing more than a two millisecond, three cubic millimeter of space-time. But an occasion could conceivably comprise an entire galactic era.

Given this, let us reconsider your discussion with Smith and Jones. The statement “on that occasion, there was no *one* person whom Smith described as a cad” is true, if by “that occasion” you mean the whole discussion. But there were many occasions *within* the discussion. There was the time Jones spilled his beer; the time Smith made an unseemly pass at the waitress; the awkward moment when you realized that you hadn’t brought any money with you; the time Smith used the name “Yaroblinsky”. Within that discussion Smith described many people as cads. But *given any one of those people*, there was indeed some occasion, within the discussion in question, such that, on that specific occasion, Smith described that person, and *only* that person, as a “cad”. Let x be *any* one of the unfortunates who, in that conversation, Smith described as a cad. Within that conversation, there was some occasion such that, on that occasion, Smith described x, and x alone, as a cad. Of course, Smith’s referring to x as a cad involved Smith’s using some name: “Argo” or “Peewee” or “Ichabod” or “Beezer”. Smith had to say “*Argo* is a cad” or “*Beezer* is a cad”. He couldn’t just say “__ is a cad” or “x is a cad”.

Let N be *any* one of the names Smith used, in the discussion in question, to ascribe to somebody the property of being a cad; and, as before, let x be *any* of the people to whom Smith ascribed that property. Within the conversation in question, there was some occasion such that, on that occasion, Smith described x, and x alone, as a cad and such that Smith used N to refer to Smith. Thus, within the conversation in question, there was some occasion such that, on that occasion, Smith described some person x, and x alone, as a cad and such that Smith used "Argo" to refer to x. Within that same conversation, there was some *other* occasion such that, on *that* occasion, Smith described some person x, and x alone, as a cad and such that Smith used "Peewee" to refer to x. Exactly the same thing (*mutatis mutandis*) is true of "Ichabod" and "Beezer", and each of the other aforementioned names.

Let us focus on "Argo". You are a part of the conversation in question; in any case, you overhear it. Within that conversation, there was some specific occasion on which Smith described some *single* person x as a cad; moreover, Smith referred to x as "Argo". Here, of course, the term "occasion" is covering a small window of time. But, uncontroversially, *within* that narrow window, there was some *one* person whom Smith described as a cad, and to whom Smith referred as "Argo". Since you are part of the conversation, you know this. When you hear Smith say "Argo is a cad", you know that Smith is referring to a single person; and you can distinguish the occasion – the narrow window of time – in which Smith produced that sentence from the narrow window of time in which Smith said "Peewee is a cad". And you can distinguish both occasions from the narrow window of time in which Smith aid "Beezer is a cad". You can cognitively keep all of these occasions separate. Given any one of these occasions, you know that Smith was describing exactly one person x as a cad; and you also know which name Smith used to make that statement. So there is some specific occasion O such

that, on O, Smith describing a single person x as a cad, and Smith used "Argo" to refer to x. And you know this: you know it merely by virtue of being a part of the colloquy in question. So merely by virtue of being a part of that discussion – nay, merely by virtue of over-hearing it – you know the truth of some claim of the form:

(*) *on such and such occasion, somebody x was uniquely such that Smith described x as a cad, and Smith used "Argo" to refer to x.*

There is only a trivial difference between (*) and:

(**) *on such and such occasion, somebody x was uniquely such that Smith described x as a cad, and "Argo" refers to x.*

So *merely* by virtue of being in a position to "pick up" use of the term "Argo", you are *ipso facto* in a position to know the truth of (**). Thus *merely* by virtue of being in a position to "pick up" use of the term "Argo some existence claim of the form: there is some x such that x uniquely has phi and "Argo" refers to x.

It is plain that what *enables* you to pick up the term "Argo" is you knowledge of some claim of the sort just described. Your picking up a term like "Argo" -- i.e. your appropriating it in manner (iii), as opposed to manner (ii) or (i) – involves your witnessing a sentence of the form "…Argo…" It involves your encountering a case of somebody's ascribing a property to Argo. (This is true even if you read such sentence: you are *still*, in effect, encountering somebody's using that sentence to make such a statement.)

Whenever you witness a token of “Argo has phi” you are witnessing an occasion such that, on that occasion, there is somebody x such that, *on that occasion*, the property of being a phi is being ascribed to x and x alone; and you are also witness that, on said occasion, “Argo” is being used to refer to x. So if you “pick up” use of the word “Argo”, that involves the following: there is some specific occasion O such that you know that, on that occasion, exactly one person x is being described as having phi and “Argo” refers to x.

We must be careful to distinguish the following claims:

(a) Exactly one person has phi.
(b) Exactly one person is being *described*, on some particular occasion, as having phi.

There are many cads. But there are many occasions where only one person is being *described* as a cad

Your picking up the word “Argo” doesn’t involve your learning the truth some claim like: exactly one person x is a cad, and “Argo” refers to x. Such a claim is false; there are many cads. Your picking up “Argo” involves your learning the truth of some claim: on some specific occasion, exactly one person x is being *described* as a cad, and “Argo” refers to x.

We must take care to make one other distinction. Smith *himself* never uttered an existence claim of the form: there is some x such that x uniquely has phi and “Argo” refers to x. Smith *himself* never uttered an existence claim of the form: *on occasion t, I (Smith) referred to somebody x as a cad and “Argo” refers to x.* Smith said only “Argo is a cad”, “Peewee is a cad”, and so forth. But the information that you are given, by virtue of being in discussion with Smith, is not confined to what Smith himself *says.* When Smith *says* “Argo is a cad”, the information that is encoded in his *words* is

quite threadbare. But you are given information above and beyond that information; you witness Smith's speech act. You put that speech act within a larger context; you have an enormous amount of information about when that speech act occurs. You know that it occurred after certain other events (e.g. before Smith spilled his beer) and before certain other events (e.g. before Jones made the unseemly pass at the waitress). In any case, even though what Smith was *saying* might have been minimal, the information you have *about* Smith's saying it – about the context in which that speech-act occurred -- is enormous; and *that* information does sufficient to give you the right kind of uniquely individuating description. Smith's words do not *by themselves* give you a uniquely individuating description that applies to Argo or Peewee or Ichabod. But there is something that *does* give you such a description: that something is your perception of the conversational context in which Smith uses the word "Argo".

There is no vicious circularity in this account. Smith personally met Argo. He learned the meaning of "Argo" through an ostensive definition. Maybe Smith *himself* gave Argo the name "Argo". In any case, Smith *saw* Argo, and then learned (or stipulated) that the thing he was seeing was named "Argo".

You meet Smith. You have a conversation with him. On some specific occasion within that conversation, Smith says "Argo was a cad". You thus know that, on that occasion, there somebody x such that Smith is describing x, and only x, as a cad, and "Argo" is how Smith is referring to x. the uniquely individuating description you associated with "Argo" is not: "the unique person x such that x is named 'Argo'". Rather, it is: the unique person x such that, on such and such occasion, smith was describing x as a cad."

In the circumstance just described, you are only at one remove from somebody who was directly acquainted with Argo – with somebody to whom "Argo" was defined ostensively.

But what we've said applies equally to the case where you are at five hundred removes from the referent of the word in question. Consider this scenario. Smith doesn't personally know Argo. But Smith knows Brown, and Brown personally knows Argo. Smith is not given an ostensive or descriptive definition of "Argo". But Smith overhears Brown saying "Argo is a cad" or whatnot; and on this basis, Smith is able to add "Argo" to his own lexicon: we just described the principles governing such an addition. You then overhear Smith saying things like "Argo is a cad", "Argo is a mean person", or whatnot. You are not given an ostensive or descriptive definition of "Argo". But on the basis of what you overhear, you are able to add "Argo" to your lexicon: the principles that govern *Smith's* appropriation of "Argo" obviously apply to your appropriation of that name. By exactly similar reasoning, it wouldn't matter if you were five hundred removes from somebody who was actually acquainted with Argo: the same principles would apply; and in each case, what would enable to use "Argo" to refer to Argo would be your knowledge of some claim of the form: something x uniquely has phi and "Argo" refers to x (where phi would be some property of the form: *is a person who, on such and such occasion is being described as having thus and such property*).

OF course, words undergo degradation and alteration as they change hands. The first person who uses the word "Argo" might pronounce one way; the people who "pick up" that name from that person might pronounce it differently. But this fact obviously doesn't affect the principles governing appropriation of that word. Brown might pronounce "Argo" one way; Smith might pronounce it differently; you might pronounce it a third way; the people to whom *you* transmit this word might pronounce it a fourth way. After a few transmissions, what is left might bear extremely little resemblance to Brown's pronunciation. In some contexts, that fact is of

importance. But that fact has no bearing on *our* point: that, whenever somebody "picks up" a name – whenever they add it to their lexicon in manner (ii) – it is in virtue of their having knowledge of a uniquely individuating description that applies to the referent. It is obvious how these points apply to names like "Socrates". The word that Socrates' own acquaintances used to refer to him was not "Socrates"; that word probably had some phonetic similarities to "Socrates"; but there were also non-trivial differences. Nonetheless, however distorted "Socrates" may be relative to the word that the Ancient Greeks used, the principles governing the "picking up" of that word – or any one of its ancestors or descendants, no matter how much phonetic distortion is involved – are the same. In each case, whenever one adds it to their lexicon in manner (ii) – it is in virtue of their having knowledge of a uniquely individuating description that applies to the referent, namely Socrates.

So suppose that "Sukrat" is the Ancient Greek word for Socrates. Glaucon learns what "Sukrat" means through an ostensive definition. Glaucon is engaged in a discussion with Zenon. Glaucon never gives Zenon any ostensive or descriptive definition of "Suktrat". But Glaucon *does* produce a sentence of the form "…Sukrat…" Thus, for reasons discussed, Zenon acquires the ability to use "Sukrat" to refer to Socrates; and, for reasons discussed, Zenon's ability to this involves his knowing a uniquely individuating description that applies to Socrates. Zenon has a conversation with Phoebus. Zenon slightly mispronounces "Sukrat". In a manner exactly analogous to that just described, Phoebus acquires the ability to use that, slightly mispronounced version of "Sukrat" to refer to Socrates. In this way, variants of "Sukrat" are transmitted from person to person, from generation to generation. What we are left with, two thousand years later, is a word that is quite different (at least phonetically) from "Socrates". But one's acquiring

the ability to use that word, however distorted it might be relative to the original word for “Socrates”, involves one’s learning some uniquely individuating description that applies to Socrates; it involves one’s learning some true claim of the form: somebody x uniquely has phi and “Socrates” has x. But, in most cases, “phi” will be some property like *is described, on such and such occasion, by so and so, as having thus and such property.* “Phi” will *not* be a property of historical moment; it will *not* be a property like *great philosopher of antiquity who drank hemlock* or *philosopher who figured as the protagonist in most of Plato’s dialogues.* The property in question will be some highly contextual one, some property that doesn’t connect in any significant way with Socrates’ intellectual or historical countenance.

Let us take stock. There are three ways you can add a referring term to your lexicon:

(i) you can be given an ostensive definition; or
(ii) you can be given a descriptive definition; or
(iii) you can witness use of that word, and simply “inherit” it.

We’ve found that whenever (i) or (iii) is the case, your ability to use the word involves your having knowledge of a uniquely individuating description of the referent of that word. Let us now consider the second of these three categories: descriptive definition.

Here it is very easy to fall prey to a confusion. Imagine the following. You have never heard the word “Socrates” before. You are talking with your friend Charlie; he says “Seymour is a latter day Socrates”. You don’t know who “Socrates” refers to? So you ask: “Who is Socrates?” Charlie says: “Socrates was the author of the Nichomachean Ethics and the Posterior Analytics”.

Of course, Charlie is wrong; Socrates did *not* write either of those works; and "Socrates" does not refer to the author of either of those works. At the same time, in consequence of your having this exchange with Charlie, you can use the word "Socrates" to refer to Socrates.

Here it *seems* as though you are able to refer to use the word "Socrates" to refer to Socrates, even though your only belief about him is wrong – even though the only thing you "know" about him isn't true. But this appearance is misleading. What we have here is a case where you add a word to your lexicon in manner (iii). You *do* know some uniquely individuating description that applies to Socrates. But that description is not "the unique author of the Nichomachean Ethics and the Posterior Analytics". Rather, it is: *the person to whom Charlie, on occasion *t*, ascribed the property of writing the Nichomachean Ethics and the Posterior Analytics*[435], for some value of *t*.

In most cases where we *seem* to be appropriating a name in manner (ii), we are *in fact* appropriating it in manner (iii). But we can certainly *construct* a case where, unmistakably, we are appropriating a name in manner (iii). (Here I basically repeating a point made, in different forms, by Kripke, Kaplan, and Evans.) Suppose you and Charlie are having a conversation. Neither of you has any idea who invented the zipper. You and Charlie decide that, at least when speaking to each other, you will refer to that person, whoever it is, as "Julius". So you have no idea who invented the zipper. But you kow that somebody did, and you have stipulated that Julius refers to that person. Thus you have added a semantic rule to English (or, at any rate, to a certain idiolect of English). It must be made clear what that rule is. The rule is given by the following:

(Z) Somebody x uniquely invented the zipper, and "Julius" refers to x.

The descriptive information is given *wide*-scope with respect to the "refers to" operator. So if Smith was a unique zipper-inventor, then "Julius was a millionaire" means *Smith was a millionaire*. That sentence does *not* mean: *somebody x uniquely invented the zipper and x was a millionaire.* "Julius" is a name, not a quantifier.

You and Charlie don't know who "Julius" refers to. (nether do I.) But it does refer to somebody. On the face of it, it certainly *seems* as though you can use it significantly: you can, with understanding, say things like "Julius was probably a talented inventor" and "Julius must be worth a fortune".

Some people have said that, under the circumstances, you *cannot* say things like "Julius must be worth a fortune" and really *mean* them. This may or may not be true (I will weigh in on this shortly). But what everybody will agree on is this. *If* you and I can use the term "Julius" with understanding, it is only because we know some true claim of the form: somebody x uniquely has phi, and "Julius" refers to x. Here, of course, phi is the property of inventing the zipper.

I do not wish to be misunderstood. A strong case can be made that unless you know *who* invented the zipper, you don't know which proposition is semantically encoded in "Julius must be worth a fortune", and other such sentences. What you know, it could be argued, is some like: *for some object O*, the sentence "Julius must be worth a fortune" is true exactly if O must be worth a fortune. Since you don't know which object invented the zipper, you don't know who "Julius" refers to; and you therefore don't know which proposition is encoded in that sentence. But, at the same time, you have a kind of metalinguistic knowledge *about* "Julius must be worth a fortune" which, in some respects, simulates real understanding of it.

Be all of this as it may, what is clear is that *if*, under the circumstances

described, you or Charlie understand "Julius must be worth a fortune" – *if*, under those circumstances, you understand "Julius" – it is entirely because you know the truth of some claim of the form: somebody x uniquely has phi [the property of inventing the zipper], and "Julius" refers to x.

In everyday life, there are many cases where we *seem* to be adding a name to our lexicon in manner (ii). Suppose little Timmy has never heard the name "Shakespeare" before. You say to him: "Shakespeare is the person who wrote *Romeo and Juliet*". Here it *seems* that the three year old is appropriating the name "Shakespeare" in manner (ii); for what you are telling him is, in effect: *there is some x such that x uniquely wrote Romeo and Juliet and "Shakespeare" refers to x.*

But really little Timmy has appropriated "Shakespeare" in manner (iii). A Kripke-style argument shows this. After your conversation with him, little Timmy can make statements about Shakespeare; they may be wrong and ill-informed; but they are wrong and ill-informed statements *about* Shakespeare. Let us suppose he does so; he asks "did Shakespeare like French fries?", and says "Shakespeare must have been weird, because his writing is so funny".

But what if it turns out Shakespeare did *not* write *Romeo and Juliet*? What if it turns out that, although he wrote *Hamlet* and *King Lear* and *Timon of Athens*, he didn't write Romeo and Juliet? Does that mean that little Timmy's old statements were blanks? Was the three year old not referring to anybody? It seems not; it seems that little Timmy did manage to ask whether *Shakespeare* liked French fries.

But if Little Timmy's ability to refer to Shakespeare depended on Shakespeare's having written *Romeo and Juliet* – if little Timmy's ability to refer to Shakespeare depended on the accuracy of the statement: *somebody uniquely wrote Romeo and Juliet and "Shakespeare" refers to x* – then his statements would indeed have been abortive.

What is going on here is that little Timmy added “Shakespeare” to his lexicon in manner (iii). He heard *you* refer to him; so the existence claim through which he added “Shakespeare” to his lexicon is something like: *there is some x such that, on such and such occasion, Daddy said that x wrote Romeo and Juliet and “Shakespeare” refers to x.* And *that* existence-claim remains true, even though the other one -- *somebody uniquely wrote Romeo and Juliet and “Shakespeare” refers to x* – turns out false.

What role did this other existence claim – the one that turned out wrong – have? Suppose you see meet somebody at a party; you don’t know him; but he acts as though he is very important. You ask your companion “who is that guy? What gives him the right to be so self-important?” You are told: “he is the dean”. Before you know who that person was, you were able to refer to him, and also to have thought about him. You were able to think *he is obnoxious* and ask “what gives him the right to act in such haughty manner?” But, despite this fact, there is still *some* sense in which you don’t know who he is. You can be able to think about, and refer to, somebody, without knowing quite *who* they are. When you tell little Timmy “Shakespeare is the guy who wrote Romeo and Juliet”, you are telling him (or trying to tell him) *who* Shakespeare was: and this, as we’ve just seen, is different from transmitting to Timmy the ability to think about, or refer to Shakespeare. (Of course, what we’ve said about Shakespeare and “Shakespeare” applies to inanimate objects and the words that label them.) Most “descriptive definitions” are really cases where one is learning *who* somebody is, or *which* object something is; they are *not* cases where one is acquiring the ability to refer to, or think about, that thing. Usually, the ability to refer to, and think about, the object in question pre-exists

There *are* cases where one is genuinely acquiring a name in manner (ii); but they are not easy to produce; and they tend to be somewhat artificial.

But here there is only one really important point. *Whenever* one adds a referring term to one's lexicon, that involves one's knowing some uniquely individuating description of the referent: for any name N, then one's adding N to one's lexicon involves one's knowing some true claim of the form: *somebody x uniquely has phi, and N names x*. (This is true, we found, even in the case where one acquires the name through ostensive definition.) The existence claims usually are not ones of historical moment; they are not ones like *somebody x wrote Romeo and Juliet and "Shakespeare" names x.* They are usually highly specific to the peculiarities of one's own situation. But what is important is that knowledge of *some* such claim, historically trivial though it may be, is needed to use a named to refer to a thing. I wish to be clear on one point. Even though we learn what a name refers to *through* some existential-descriptive claim, the name itself is just a label. Semantically, "Shakespeare" simply labels Shakespeare; it doesn't have any descriptive content (except such as it has in virtue of being a mere label of Shakespeare). But one always learns *which* thing a name labels *through* some existential-descriptive claim. The semantics are not existential-descriptive; but the act of *learning* the semantics is existential-descriptive. The semantic rule for "Shakespeare" must be *given* to you in some way or other. But what is given is not the same thing as *how* it is given. The following statement is false:

(S1) The semantic rule for "Shakespeare" is this: somebody x uniquely wrote *Hamlet* and "Shakespeare" names x.

S1 is false because "Shakespeare" refers to Shakespeare in worlds where *Hamlet* was never written. So the semantics for "Shakespeare" cannot have anything to do with *Hamlet.*

But the *following* statement is true:

(S2) Somebody x uniquely wrote *Hamlet* and, and the semantic rule for "Shakespeare" is this: "Shakespeare" names x.

The operator "the semantic rule for__ is" must be given *narrow*-scope. Otherwise we end up committed to a false descriptivism. The name is just a label. The literal meaning of "Shakespeare was prolific" is simply: *Shakespeare wrote Hamlet.* It is *not*: somebody x uniquely wrote Hamlet and x was prolific.

Though false, S1 is not, so to speak, *as* false as the following (the asterices are to be taken as quasi-quotation marks, i.e. as being just like quotation-marks *except* that they don't quote any variables that they enclose):

(S3) The semantic rule for "Shakespeare is this: for any predicate *…x…*, *…Shakespeare…* means: somebody O uniquely wrote Hamlet and…O…

What is literally meant by "Shakespeare was bald" is simply *Shakespeare was bald*; it is not: *somebody x* uniquely wrote Hamlet and x was bald.

To avoid running afoul of Kripke's correct points, we must take care to give the relevant operators – "refers to", "is the semantic rule for__" – *narrow* scope.

Another point is worth bringing up. One typically forgets the existence-claim through which one added a name to one's lexicon. One typically forgets the *fist* time one heard "Shakespeare" or "Aristotle"; one often forgets one's *first* meeting of a person. Given this, one is tempted to say:

"You don't necessarily associate *any* uniquely individuating description with "Aristotle" or "Smith" or "Jones". Such a description was needed to

enable you to refer to those people, and to think about them. But once that ability is in place, the description drops out. At that point, there can be *pure* thought about the entities in question. You can *just* think about Aristotle and Smith; you don't have to think about them *through* a uniquely individuating description.

This is not how it works. What happens is not, I propose, that one *ceases* to have knowledge of a uniquely individuating description, but rather that one replaces one such description with another one: a related, but distinct one.

Remember what we said earlier about "Fred". If you add "Fred" to your lexicon in manner (i), i.e. through an ostensive definition, that involves your knowing some claim of the form: *somebody x uniquely has phi and "Fred" names x.*

Suppose you and a friend are at a party. You see somebody wearing a bowler hat, talking animatedly. At time t, your friend points to that person and say "*that* is Fred". Your adding "Fred" to your lexicon involves your knowing some existence claim, something like:

(e1) *there is now somebody x whom I am sense-perceiving and attending to, and who is wearing a bowler hat, and "Fred" names x.*

Let us suppose that, a minute or so after "Fred" is ostensively defined for you, you go up to him and start talking. Now you have new information about Fred; you have a new uniquely individuating description of him. There are a number of reasons for this. First of all, the mere lapse in time between your current situation – your now talking with Fred – and the initial ostensive definition by itself guarantees that you will have a new uniquely individuating description of him. At time t, (e1) was your uniquely

individuating description. But now that time has passed, (e1) has been replaced. By itself, the mere passage of time replaces (e1) with:

(e2) *there is somebody x whom, a minute or two ago, I was sense-perceiving and attending to, and who was then wearing a bowler hat, and "Fred" names x.*

notice that (e2) "interlocks" with (e1). The thing having the property specified in (e2) is the same as the thing having the property specified in (e1). The uniquely individuating description changes. But the new one is related to the old one in such a way that something x uniquely makes both of them true. The new one "inherits" the verifier of the old one. In a moment we will say more about this process of inheritance or transmission.

But other things, besides the mere passage of time, will replace (e1), with a new uniquely individuating description. You are now engaged in conversation with Fred. You can see him more clearly (although from afar he looked hale and strong, you now see that he is puffy and out of shape); he is saying certain things to you (he is talking about the national debt). As a result, you generate a *new* uniquely individuating description of Fred. Your new one is something like:

(e3) *there is somebody x whom, a minute or two ago, I was sense-perceiving and attending to, and who was then wearing a bowler hat; and I am not attending to x, and conversing with him, x is now talking about the national debt, and x looks puffy and unhealthy, and "Fred" names x.*

Notice that (e3) and (e2) interlock; (e2) transmits its verifier to (e3).

Now that you have (e3), you can *drop* (e2) and (e1). You can completely forget them; and that will not jeopardize your ability to think about. Suppose a moment passes; you have stopped talking with Fred. By itself, the mere passage of time replaces (e3) with a slightly different claim:

(e4) there is some x such that, a moment ago, x was talking about the national debt, and x looked puffy and unhealthy, and "Fred" names x.

Suppose that you have completely forgotten (e1)-(e3). You *still* have some uniquely individuating description of Fred. And, so long as you are able to think about Fred, this will continue to be the cases. You go home at night. (e4) is now replaced with:

(e5) there is some x such that, several hours ago, at a party, x was talking about the national debt, and x looked puffy and unhealthy, and "Fred" names x.

This process can continue *ad infinitum*. One's ability to think about, and refer to, something can only be *started up* by one's acquiring knowledge of the right kind of existence claim. Typically, that specific claim is forgotten. But it is replaced by another "interlocking" existence claim – another interlocking uniquely individuating description. And that second one will, in its turn, be replaced by a third; and that third by a fourth, and so on.

The point is that the ability to think about, and refer to, an external object always involves possession of a uniquely individuating description that applies to it.

None of this belittles the role that causal connections play in conception

and reference. I am able to think about, and refer to, Fred precisely because I *saw* him: and seeing something involves being causally connected to it. But *how* is the causal connection involved? What enables me to think about Fred is my knowledge of some uniquely individuating description: *there is some guy whom I am now seeing and attending to, who is wearing a bowler hat, and "Fred" names that guy.* The causal connection is involved in my seeing Fred; and my seeing Fred, in its turn, is involved in my having a uniquely individuating description of him. So what is doing the work is my knowledge of a uniquely individuating description of Fred; the causal connection, though essential, has a role that is sub-ordinate and internal to that my knowledge of that description.

Here we need to very clear about a certain distinction: the distinction between *literal meaning* and *cognitive content.* "Fred" is just a label. The following statement is true:

(F1) There is somebody x who I met on such and such occasion, and who looked pale and talked about communism; and the semantic rule for "Fred" is: "Fred" names x.

The following statement is false:

(F2) The semantic rule for "Fred" is this: There is somebody x who I met on such and such occasion, and who looked pale and talked about communism; and "Fred" names x.

The following is even *more* false

(F3) The semantic rule for “Fred” is this: for any predicate *…x…*, *…Fred…* means: There is somebody O who I met on such and such occasion, and who looked pale and talked about communism and…O…

The literal meaning of

(*) “Fred looked puffy and pale”

is simply:

(*LM) *Fred looked puffy and pale.*

But remember that one knows what “Fred” refers to only in so far as one knows some claim of the form: *somebody x uniquely has phi and “Fred” refers to x.*

---

[1] If there were no sentient beings, it would still be true that triangles have three sides. Therefore it would still be a truth that that triangles have three sides. Therefore it would still be a true proposition that that triangles have three sides. If we try to say that truths and falsehoods come into existence with the onset of awareness, we must say that – even though there are no truths *per se* in a world devoid of mind – nonetheless such a world provides the *grounds* for truth. Though there were no truths a billion years, the *grounds* were still present for the truth that there were a certain number of planets. That is why, when a mind enters such a world, it finds itself surround by objective truths.

But to say that the *grounds* for truth are present is just to say that the truths themselves are present. So I don't see that the foregoing argument eliminates the existence of mind-independent truth.

[2] Quine (1990: 26).

[3] Richard Montague (1974) showed how to convert any sentence into one of subject-predicate form; arguably, he thus showed that any proposition is given by a sentence of that form. So the trouble with Kant's definition isn't exactly that it presupposes that all sentences (or propositions) are of that form. But after we use Montague's method to convert (say) "if P, then P" into subject-predicate form, Kant's definition of analyticity turns out to be false. After the conversion, that sentence becomes "if* <P, P>", where "if* denotes a function assigning truth to an ordered pair of propositions exactly if the second proposition is a consequence of the first. So the subject becomes the ordered pair <P, P>, and the predicate becomes the second-level function denoted by "if*". But this doesn't help Kant. There is no sense, not even a metaphorical one, in which <P,P> is "contained" in the concept denoted by "if*".

[4] Strictly speaking, *A=A* is not a proposition. What *is* a proposition is *for all a, A=A* or *Smith=Smith.* When I talk about "A=A", I mean some specific proposition of that form or the proposition that all propositions of that form are true.

[5] Kant (1965: 52)

[6] See Pap (1958).

[7] Kant (1965: 52-54)

[8] Kant (1965: 52-54)

[9] Kant 1965 (53)

[10] See Bonjour (1998: chapter 1).

[11] Reichenbach (1958) denies this assumption. He thinks that the data of sense-perception are neutral between Euclidean and non-Euclidean geometries and that, so far as we think otherwise, we are guilty of smuggling theoretical commitments into our descriptions of our sense-experiences. I agree with Reichenbach.

[12] Bonjour (1998: 23) makes this important point.

[13] This point is made very clearly in Bonjour (1998: 25).

[14] See Bonjour (1998: 20-26).

[15] Arguably, a talented mathematician wouldn't *have* to base his answer on this kind of play of imagery. Instead he might be able to use his knowledge of correspondences between geometry and number-theory: just as number-theoretic truths can be learned through spatial intuition, so perhaps geometrical truths could, in principle, be learned through number-theoretic intuition. But this only reinforces my point: usage of imagery is a mere *vehicle* for some kind of conceptual analysis.

For the record, I myself *don't* think that geometrical reason could ever be completely "arithmetized". I don't mean to deny the existence of any objective correspondence between those two disciplines. I mean only that, if one is really thinking about geometry, one eventually has to relate one's results to *shapes*. Even though shapes can be represented by sets of numbers, one isn't thinking of a thing *as* a shape if one thinks of it as a set of numbers. Geometrical thought eventually involves thinking of something *as* a shape. So, in my view, geometrical thought does ineliminably involve *some* component of visual intuition (though given enough mathematical acumen, there is no limit to how marginal that component can be).

At the same time, even in geometrical thought, the element of visual intuition is still a vehicle for conceptual analysis; it is a way of analyzing the relevant geometrical concepts. It doesn't have the quasi-empirical role that Kant attributed to it. So even though Kant may have been right to insist that geometrical thought necessarily involves visual imagery, the conclusion he drew from that – geometry teaches us about our forms of thought, not about objective relations among concepts -- was quite wrong.

[16] Quoted in Burge (2000: 32).

[17] Peacocke (2000: 255).

[18] Of course, what cannot be *per*ceived can sometimes be *con*ceived: e.g. sub-atomic particles that do not reflect light. But if something cannot be *con*ceived, then surely it cannot be *per*ceived. The conditions something has to meet to be perceptible are *stronger* than those it has to meet to be *con*ceivable.

[19] Here when we talk about necessity, we are focusing on epistemic necessity – that is the only kind Kant recognizes.

[20] I develop this idea in another publication. See Kuczynski 1999.

[21] See Kuczynski 1999 for a version of this argument. I did not know it at the time, but I there put forth a "transcendental", Kant-style argument against a Humean analysis of causation.

[22] '[O]nly communities, not individuals, can be interpreted as having original intentionality.' [Brandom (1998), p. 61].

[23] Actually, Brandom holds that for a creature to have *any kind of intentionality at all*, it must first know a language: '[My view] deserves...to be called a *linguistic* [Brandom's italics] view of intentionality...because linguistic practice is nonetheless accorded a certain kind f explanatory priority over rational agency. The intentionality of non-linguistic creatures is presented as dependent on, and in a specific sense derivative from, that of their linguistically qualified interpreters, who as a community exhibit a non-derivative, original intentionality.' [Brandom (1998), p. 152].

This point implies a vicious regress. Knowing a language is itself a form of intentionality. So if intentionality could be correctly attributed only by some creature that already had intentionality, then that creature, *qua* creature possessed of intentionality, could exist only if some other language speaker existed. But then that *that* language speaker, *qua* creature possessed of intentionality couldn't exist unless some third language speaker existed; and so on *ad infinitum.* There couldn't be a single intentional creature unless there were an infinity of them.

Also Brandom's point is false on purely empirical grounds. Dogs and infants obviously have intentional mental states. Finally, Brandom's own account of how creatures learn language clearly presupposes that pre-linguistic creatures have intentionality of some kind. According to Brandom, creatures learn a language by entering into communal practices. But how a could a non-intentional creature enter into communal practices? (Of course, a non-intentional robot could simulate being engaged in communal practices. But it couldn't really have entered into such practices.) Brandom's position begins to make sense only if we take him to mean *cognitive intentionality* by 'intentionality'. (By 'cognitive intentionality' I mean the kind of intentionality one has in being aware of propositions).

[24] One error in this argument is the idea that anything mediating information is *ipso facto* a language. That thesis obliterates the distinction between distinctly *linguistic* representation and non-linguistic (e.g. perceptual) representation, and thus trivializes the idea that thought is symbolic.

[25] John McDowell holds this view. [Searle McDowell (1994) and (1998)]

[26] Searle gives this argument. [Searle (1992), and (1999)]

[27] This is of course in direct contradiction to what Jerry Fodor holds. [Fodor (1975)]

[28] At this point some important disclaimers must be made. Wittgenstein is notoriously difficult to interpret. It is a matter of special controversy how to interpret his 'rule-following' and 'private language' arguments. And it is with these arguments specifically that a great deal of this paper will be concerned. To say the least, I cannot guarantee that my interpretations of Wittgenstein are correct. [28] (They are, I think, fairly orthodox criticisms — at any rate, they are orthodox relative to a certain tradition of Wittgenstein scholarship.) Given this alone, it might seem as though my whole project is in peril. But this is not so. Although I cannot be sure that my interpretations of Wittgenstein are correct, I do think that my interpretations of Wittgenstein correspond very well to Brandom's interpretations of him. (I submit that a close reading of *Making it Explicit* will quite bear this out.) Given that, in so far as I am concerned to refute any author, it is Brandom and not Wittgenstein, it is enough if we have a grip on how *Brandom* understands Wittgenstein. Also, even if (what I have some confidence is not the case) we have misunderstood how Brandom misunderstands Wittgenstein, that would not still invalidate our project: for the *main* purpose of that project — over and above refuting Brandom — is to show that the ability to learn a language presupposes the ability to grasp propositions. Our case for this position will not presuppose the correctness of this or that interpretation of Brandom or of Wittgenstein. (Nonetheless it is to be hoped that I have understood Brandom correctly since I frame many of my arguments for cognitive realism in terms of what I have taken to be Brandom's criticism of it; and my

arguments would thus lose a certain *rapport* with current discussion if I have mis-understood Brandom.) In any case, readers who are worried about Wittgenstein's views being mis-represented, might do well to mentally replace each occurrence of 'Wittgenstein' with an occurrence of 'Wittgenstein$_B$' or 'Wittgenstein$_K$', where the former means 'Wittgenstein *as understood by Brandom*' and the latter means 'Wittgenstein *as understood by the author*'.

[29] In the *Blue Book* Wittgenstein says that 'thinking consists in operating with symbols'. (Thinking may surely be identified with seeing relations among propositions.) In the *Philosophical Investigations* W. says that one cannot know how to operate with symbols unless one is a member of a community [Wittgenstein (1999) sections 143-279, especially sections 202 and 208)].

[30] This is not my own view. This is what a pro-Wittgensteinian would say. I happen to think that all genuine abilities - all cases of knowing how - involve a grasp (though usually an unconscious one) of propositions.

[31] I have just set forth a rather liberal, but (I think) basically accurate, reconstruction of sections 139-271.

[32] Of course, if you react to the sound 'fire' by *understanding* it, then trivially this kind of reaction does betoken understanding. But Wittgenstein-Brandom is trying to explain what understanding is in terms of pre-cognitive dispositions to react.

[33] Actually, brains *aren't* conscious. Creatures which *have* brains are conscious. ('Brains don't think. *People* think.' *Language, Sense, and Non-sense*' P.M.S. Hacker and Gordon Baker.) But, for the sake of argument, we will ignore this point. Anyway, it isn't really essential to the interlocutor's point.

[34] Wittgenstein made this very point in the *Philosophical Investigations* [p. 174].

[35] Direct reference theory seems at first to deny this. But I'm not sure that ultimately it really does. In light of what Kripke has said, it can no longer be maintained that 'Socrates' is equivalent to some definite description. But does it follow that 'Socrates' expresses no *concept* whatsoever?

[36] How could one give and ask for reasons without previously grasping reasons? Of course, Brandom has a reason for saying this: one must engage in communal, linguistic practices to grasp propositions. So - paradoxical though it may sound - giving and asking for reasons is prior to grasping reasons. Well, we have seen that this central contention of Brandom's is false and that it *really* is a paradoxical assertion, as it first appears. Nevertheless, with a slight amendment, what Brandom is saying here is basically correct. What Brandom must say here, if his analysis is to be correct, is - *not* that a human being, whereas a parrot cannot, 'given and ask' for reasons - but simply that a human being can, whereas a parrot cannot, grasp reasons at all. I am ignoring this odd feature of Brandom's explanation. Thus amended, Brandom's analysis is correct.

[37] Cf. Brandom: 'The use of sentences is prior in the order of explanation to the use of subsentential expresses because sentences are the only expressions whose utterance "makes a move in the language game."' [Brandom (1998), p. 82] 'Referring to something, indicating or naming it, is also something one do with linguistic expressions; it is a speech act one can perform. But these belong to a class of speech acts that is in an important sense derivative from or parasitic on speech acts involving sentences, paradigmatically claiming, asserting, or putting forward as true. The significance of taking or treating something as a name, as purporting to refer to an object, consists in how one takes it to be proper to use the expression, and the use of the expressions as names is unintelligible except in the contest of using expressions containing them as sentences.' [Brandom (1998), p. 82].

[38] Of course, given a non-assertoric sentence containing 'Bill Clinton' — e.g. 'Does Bill Clinton

smoke?' and 'Make Bill Clinton come home' — the term 'Bill Clinton' refers in that sentence to Bill Clinton because, and only because, the *assertoric* sentence corresponding to that sentence is such that it is true if and only Bill Clinton has a certain property. So 'Bill Clinton' refers to Bill Clinton in 'Does Bill Clinton smoke' because, and only because, 'Bill Clinton Smokes' is true if and only Bill Clinton smokes.

[39] This argument was (basically) set forth in the *Philosophical Investigations*, sections 1-10.

[40] Note that this is a parallelism between Brandom's account of how *people* acquire intentionality and how *words* acquire meaning. It is initially *society* that has intentionality; the individual's intentionality is derivative of that of the collection. Analogously, it is originally *sentences* that have meaning; the individual word's meaning is derivative of that of the sentence's.

[41] I believe Chomsky says as much. [Chomsky (1990)]. I don't think that Putnam-Burge externalism bears against this. What Putnam-Burge externalism shows is *not* that there cannot be a set of privately grasped linguistic rules, but that the linguistic rules that we *do* privately grasp — and that are, in effect, private — sometimes make reference to the activities of other people. I can have a private language whose semantic rules are sometimes of the form *let "gold" refer to whatever it is that Charlie discovered.* A rule that makes reference to other people's thoughts and deeds may still be a private rule — may still be a rule that belongs to a private language. A 'private' language is not necessarily one that has to exist in a void of other people. It is one that one can subject *oneself* to. Obviously the semantic rules for such a language could refer to other people without ceasing to be private.

[42] Of course, all of these authors found additional reasons for believing that creatures must know a language to be able to think. But I think that they sought these reasons to rationalize an existing belief that linguistic competence underlies thought; and surely this existing belief was motivated by the obvious fact that we are so much more developed cognitively than other creatures and that we are the only creatures who really have language. My own view is that Brandom's and Wittgenstein's arguments are much too obscure to have commanded any agreement, even by their creators, had people not *already* believed that the ability to think somehow supervened on the ability to speak — had people not *already* assented to the conclusions of those arguments.

[43] Interestingly, I think that Freud, of all people, said this in *The Ego and the Id.*

[44] If I am not mistaken, Freud says this in 'The Unconscious.'

[45] Of course, the classic proponent of this view is Hume. Hume maintains that moral sentences are neither true nor false, and that, in believing them to be true or false, we are guilty of a kind of delusion. As human beings, we tend to project our feelings onto the world and mistake these feelings for objective realities. Moral sentences embody these projective delusions. (Hume [1], Book II; Hume [2].) Simon Blackburn defends a Humean view of morality in Blackburn [1], chapter 6; and in the entirety of Blackburn [2].

In *Language, Truth, and Logic* Ayer advocates emotivism. It doesn't occur to Ayer to consider the possibility that *because* ethical judgments express emotion, they are true or false. Ayer takes it for granted the opposite inference holds: because ethical judgments express emotion, they are *not* true or false.

[46] Inevitably, some will say that this point presupposes the correctness of a "speech-act" conception of semantics, as opposed to a tractarian conception.

This is completely erroneous; my point above doesn't presuppose such a conception.

It would be nonsense to say:

"snow is white" bears the proposition *snow is white.* But if you want to express the belief that snow is white, then you'd better use some other locution; for "snow is white" won't do the trick.

No if "snow is white" bears the proposition *snow is white* then *ipso facto* it enables one to express the corresponding belief.

And if one "snow is white" enables one to express that belief, then *ipso facto* it (or uses of it) bear that proposition. Suppose someone said:

If you want to express the belief that snow is white, then use the locution "snow is white". But if you want some locution that bears the proposition, then you'd better look elsewhere: "snow is white" isn't what you want.

Again, that would be patent nonsense. In so far as tokens of "snow is white" are suitable vehicles for expressions of the *belief* that snow is white, they *bear* that proposition.

So it is surely not too far off the mark to say that "snow is white" bears the proposition *snow is white* in virtue of its being an apt way of expressing that belief.

[47] This doesn't mean that whenever one makes a moral statement one is in the throes of emotion. If I say "letting children starve (when there is plenty of food to give them) is evil", I don't necessarily have to be in a fit of rage or of outrage. But emotions are not always *occurrent.* I can hate somebody for forty years and not always be in a fit of anger. The emotion shouldn't be confused with the occurrent manifestations of that emotion. Now if somebody *felt* nothing – had *no* emotion in connection with the phenomenon of children gratuitously starving – then it is a bit hard to see how that person could *sincerely* say "letting children starve is evil".

I will operate on the assumption that this point is correct. I admit that this is an arguable point. But it is, I think, reasonable enough to warrant our investigating the consequences of it. And one of those consequences, I submit, is that Emotivism is a form of moral realism.

[48] In *The Emotions: An Outline of a Theory* Sartre argues for a view that is, at the very least, similar to the view that emotions are beliefs: "emotions are…a way of apprehending the world" (Calhoun and Solomon [1]). Robert Solomon argues for a similar view: "my emotions *are* judgments I make…" (Calhoun and Solomon [1]: 316, Solomon's emphasis). Calhoun's and Sartre's arguments are not quite the same as those that I give, though there are unmistakable similarities.

Hume epitomizes the *opposing* position – the view that emotions are absolutely incapable of being, in any non-derivative sense, true or false. See Hume [1], Book II, Section III ("Of the Influencing Motives of the Will", especially page 414:"'tis impossible…that [an] emotion can be oppos'd by, or contradictory to, truth and reason…"). This is, I think, connected with his ethical anti-realism. For if you see emotions as true or false; and if you see ethical statements as embodiments of emotion; then you must see such statements as themselves being true or false – you are therefore an ethical realist.

[49] *The Quest for Reality*. Oxford: 2001

[50] There are some superficial exceptions to this. One might say:

"The word 'Socrates' represents Socrates. But that word doesn't predicate anything of Socrates. So not all representation is predicational."

Very well. But what does it mean to say that "Socrates" represents (denotes/ refers to) Socrates? It means this. Sentences of the form

"…Socrates…"

mean

…Socrates…

For “Socrates” to represent (denote) Socrates is, ultimately, for sentences containing “Socrates” to be about Socrates. “Socrates” denotes Socrates because “Socrates was bald” is made true by Socrates’ being bald, not Plato’s. Suppose that “Socrates was bald” was made true by Plato’s being bald (but that otherwise the semantics of English was the same). In that case, “Socrates” would denote Socrates (at least in that sentential context). It would be nonsense to say:

(*)“Socrates” denotes Socrates, but “Socrates was bald” says that Plato was bald – it doesn’t say anything about Socrates.

The absurdity of (*) shows that, when we have a case of a non-predicational symbol, we are really dealing with an “incomplete” symbol: something that, in some sense, is sub-symbolic.

[51] I myself have a paper forthcoming in which I argue that emotions are *not* beliefs!

[52] A person may be unhappy, and think that he is *legitimately* unhappy – he has just lost his job, his wife just left him, etc. Under such circumstances, he may well choose to take a pill or a drink. So it is, I grant, not always the case that, when a person narcotizes himself, it is to get rid of what he perceives to be in irrational emotion.

[53] People also justify their positive emotions. Suppose a person is obviously happy that something bad happened to someone. That person will, in all likelihood, try to justify that emotion: “well, he got what he deserved; he was a liar and a cheat…”

[54] Arguably, propositions can be rational or irrational – the proposition *1+1=3* is “irrational”. But when we describe a proposition “rational” or “irrational”, we mean, I think, that belief in it is rational or irrational.

In any case, the only *mental* entities that can rational or irrational are beliefs. So if emotions answer to canons of rationality, they are beliefs.

[55] I borrowed this example from Davidson.

[56] Kripke 1980 and Putnam 1975.

[57] This statement should be qualified. I am doing a bit of reconstruction. What I should really say is: ‘An argument for dualism that might have had credibility in the times before Kripke and Putnam is this….’ I don’t know if *exactly* the argument I’m about to give was given.

[58] There are some apparent counterexamples to this principle – e.g. (*) ‘the inventor of bifocals *is* identical with the first post-master general, but the former didn’t *have* to be identical with the latter.’ But Kripke showed conclusively that these apparent counter-examples are merely apparent. Actually, it was really Russell who showed this. (*) is not an identity at all; it

says merely: some one individual x had two sets of properties –x had the property of being a postmaster before anyone else and x also had the property of being an inventor of bifocals before anyone else.

[59] If indeed there are facts about the essence of *Socrates*, i.e. if indeed Socrates had an essence. Some would deny that he does. I think that Kripke was right to see his essence as bound up with his conditions of origination.

[60] It might also be said that properties are invariant from world to world; the property of being an even number is the same world 5 as it is in world 37. This is true: properties *are* world-invariant. But necessary facts about properties, and their interrelations, are reflected in facts about concepts. The proposition *anything red is colored* is necessarily true because of facts about the properties of being red and colored. But this fact corresponds to a fact about the concepts *red* and *colored.* We can, I think, analyze the necessity of *anything red is colored* as a relation between the concepts *red* and *colored* or between the corresponding properties; we can talk *either* about concepts *or* about properties. I think it more convenient to talk about the concepts which express properties.

[61] Suppose that 'if x falls under C1, then x falls under C2' is necessary, and that one knows that. In that case, one knows – or can readily deduce – that 'x falls under C1 but not under C2' is not possible. And if 'x falls under C1 and not under C2' is possible, and one knows that, then one knows, or can immediately deduce, that 'if x falls under C1, x falls under C2' is not necessary.

[62] By 'constitutively involves' I mean *as a matter of principle* involves, as opposed to 'is causally dependent on'. Obviously, for me to grasp the concept *even number*, it may be *causally* necessary that I be embedded in a certain kind of environment. But it doesn't follow that my being thus embedded in a certain of environment is to any degree *constitutive* of my grasping that concept; it could just be that my being thus embedded is needed to 'trigger' or 'activate' my grasp of that concept. This is, in fact, what I hold.

[63] Unless, of course, it was previously an embodied brain, and through the mediation of that body had the appropriate causal liaisons to Socrates and

water.

[64] The classical statement of analytic functionalism is found in Lewis [1].

[65] Even extremely hardline functionalists like Fodor concede this. See the introduction to Fodor 1983.

[66] See Jackson and Pettit 2004b-2004e, especially 2004c (48-49).

[67] See Kaplan 1968, 1989a, 1989b, and Putnam 1975.

[68] See Jackson and Pettit 2004b-2004e, especially 2004b (58-62), 2004d (105-109), and the whole of 2004e.

[69] Perhaps one might distinguish *mental* content from *representational* content. One might say that Smith and twin-Smith have qualitatively identical mental contents, but different representational contents. There are two things to say. First of all, it seems at best arbitrary to separate so sharply between the mental and the representational. But suppose *arguendo* that we grant that distinction. In that case, by the argument just given, externalism strips representational content of causal power – not a palatable result.

[70] See Jackson and Pettit 2004b-2004e, especially 2004d (108-114) and 2004b (58-62).

[71] The same is true of most of their applications of that concept. Particularly impressive are their applications of it to social theory (see Jackson and Pettit 2004f) and to the concept of causal explanation itself (see Jackson and Pettit 2004g). For example, they argue (2004f) that in *some cases*, it is more informative to know a *program* cause than an actual cause, since the former is richer in counter-factual information. Thus, in some cases, a macroscopic course-grained explanation is actually *better* than a microscopic, fine-grained explanation. To my astonishment, I found myself in complete agreement with this radical point.

[72] "Psychophysical and theoretical identifications" *Australasian Journal of Philosophy* 50 pp.249-258.

[73] (*From Folk-Psychology to Cognitive Science: The Case Against Belief* Bradford Books, the MIT Press, 1983, pp. 25-26.) Stich's point is not to prove that functionalism is false. It is to prove something else entirely – that 'folk-psychology' is false. Stich wants to show that the categories in terms of which most psychological explanation is couched (*belief, desire,* and so forth) are explanatorily bankrupt.

[74] Propositions are not mental and are non-derivatively right or wrong. But propositions are not *spatiotemporal* entities; they are *abstracta.* Any *spatiotemporal* entity that can be non-derivatively described in terms of correctness is *ipso facto,* I think, mental.

[75] The inscription 'snow is white' is correct, but only in a derivative sense – derivatively of the intentions and linguistic practices that assign it meaning. In a world lacking such intentions and practices, it would be dead ink – as incapable of truth or falsity as a rock.

[76] "The Foundations of Mathematics" in *Philosophical Papers* (D.H. Mellor ed.) Cambridge: 1990.

[77] I should point out that Russell (in "Knowledge by Acquaintance and Knowledge by Description") contended that some propositions are ungraspable: any proposition that has an external object (as opposed to some description thereof) for one of its constituents is, for Russell, ungraspable. Contemporary direct reference theorists sometimes hold that there propositions we cannot grasp, for example, propositions that are object-involving with respect to future individuals.

[78] In this paper, I will use the term 'supervenes' to refer to what is sometimes called *metaphysical* supervenience, not causal – or any other – kind of supervenience. To say that x 'supervenes' on y – in the sense in which we are using this term – is to say that x is *wholly determined* by y; which is in turn, roughly, to say that there is no possible world where there is a change in x without a change in y.

[79] Intentionalism has recently experienced a surge of popularity. Consider the following three formulations (cited by Byrne: Byrne [1]):

Dretske: 'All mental facts are representational facts.' (Dretske [1]: xiii)

Lycan: 'The mind has no special properties that are not exhausted by its representational properties, along with or in combination with the functional organization of its components. (Lycan [1]: 11)

Tye: 'Phenomenal Character (or what it is like) is one and the same as a certain sort of intentional content. (Tye [1]: 137)

Tye is explicitly articulated what I am calling 'intentionalism'. Dretske and Lycan are articulating a position that is *stronger* than intentionalism, and that entails intentionalism: they are articulating the position that *all* mental facts are representational. Now *if* it is true that all mental facts are representational, then given that phenomenal facts and (non-derivatively) representational facts are obviously mental, it follows that there can be no *difference* in the phenomenal properties of a mental state of affairs without a difference in the representational properties of that state of affairs. So Dretske and Lycan are, at least by implication, endorsing what I am calling 'intentionalism.'

[80] In the contemporary literature, the term 'intentionalism' is used to denote two related, but distinct, doctrines: (i) the doctrine that *all* distinctly mental properties are representational properties; and (ii) the weaker doctrine that *phenomenal* properties are representational properties. Of course, (i) entails (ii), but not *vice versa.* So – though they are interrelated – (i) and (ii) are quite distinct doctrines.

In this essay, I will use the term 'intentionalism' to refer to (ii). Of course, if (ii) is false, it follows that (i) is false. Consequently, many of my arguments will be *directly* aimed at (ii), and therefore only *by implication* aimed at (i). But some of my arguments will be *directly* aimed at (i).

(i) and (ii) are not carefully distinguished in the literature. For example, in his important article "Intentionalism Defended", Alex Byrne (Byrne [1] p. 204) seems to identify (i) and (ii): for he explicitly says of the following statements that they are but different formulations of the same thesis:

(a) 'All mental facts are representational facts' (Dretske)

(b) 'Phenomenal character (or what it is like) is one and the same thing as a certain sort of intentional content. (Tye)

First of all, (b) seems to correspond to (ii), and (a) to (i). So, in effect, Byrne is failing to distinguish

between (i) and (ii). In any case, it is clear that (a) and (b) should not be identified; for (a) says (essentially) that *everything* mental is representational, while (b) says only that the *phenomenal* is representational.

[81][81] This objection was made to me by someone who read an earlier draft of this paper and who wishes to remain anonymous. If I am not mistaken, Gilbert Ryle advocates this conception of memory. Also, it seems to me that those who deny the existence of unconscious mental activity, e.g. John Searle, are compelled to advocate this conception of memory.

I think that many functionalists would adopt the view advocated by the objector.

[82] See Heck [1] for an excellent discussion of this issue.

[83] In this paper, I will use the terms "propositional" content to refer to the kind of representational content that can be embodied in propositions; and I will use the term "non-propositional" content to refer to any kind of representational content that cannot be embodied in propositions. But this is not the general practice. Usually, what I am calling "propositional" content is referred to as "conceptual" content, and I what I am call "non-propositional" content is referred to as "non-conceptual" content.

But my choice of nomenclature is, I think, to be preferred. This is because propositions can have things other than concepts as constituents, and therefore can have something that deserves to be call "non-conceptual" content. In any case, this is almost universally held these days. The idea is that the proposition *Socrates is bald* is *de re* about Socrates, and thus has Socrates himself – who is obviously not a concept – as a constituent. If Socrates himself can be a constituent of propositions, then some of the content of some propositions is something other than a concept, and is thus non-conceptual. In general, if propositions can have things other than concepts as their constituents – and it is practically universally agreed that they can – then propositions can have things other than concepts for their constituents and thus have something that might legitimately called "non-conceptual" content. So there is certainly some reason to believe that the kind of content that can be embodied in propositions is sometimes non-conceptual in nature.

Consequently, the distinction between propositional and non-propositional content is different from that between conceptual and non-conceptual content. In fact, the people who use the terms 'conceptual content' and 'non-conceptual content' to mark the distinction *I* am talking about – the distinction between, as I am putting it, propositional and non-propositional content – almost always hold that there *are* object-involving propositions, i.e. that non-concepts – e.g. Socrates, Plato – can be constituents of propositions. Given this, it is odd that, when they are discussing the question whether all representational content can be embodied in *propositions*, they often state this question as 'is all representational content *conceptual*?' Again, it seems to me that (often, though probably not always) what is *really* being asked is 'is all representational content *propositional*'' (or 'is all representational content capable of being embodied in a proposition?') and not 'is all representational content *conceptual*?

I should make one absolutely clear. In this paper, whenever I say 'non-propositional content', I mean non-propositional *representational* content. I mean the kind of content that is *representational* but that cannot be embodied in a proposition. I will question whether this kind of content exists. But, just to be clear, my term for this kind of content (if it exists) will be simply 'non-propositional content' – the term 'representational' is to be taken as implicit.

[84] The idea that there is 'non-propositional content' (usually the term 'non-conceptual content' is used – see last footnote) is very widely held. Evans, Peacocke, Heck, and Michael Tye are some of the prominent believers in non-propositional content. (See Evans [1], Peacocke [1], Heck [1], and Tye [1] and [2].) Peacocke is not himself an intentionalist (see Peacocke [2]). Michael Tye *is* an intentionalist.

(See Tye [1] and [2].) Evans did not weigh in on this particular issue and so far as I know, neither has Heck.

[85] I would like to make one point absolutely clear. Throughout this paper, by 'non-propositional content', I mean 'non-propositional *representational* content.' So the term 'non-propositional content', as I am using it, is really just an abbreviation for 'non-propositional representational content.' So the term 'non-propositional content' will be our term to refer to that kind of representational content that cannot be embodied in propositions.

Later on, I will argue that there is no such thing as representational content that cannot be embodied in propositions. But if only for discursive reasons, we must have a term to denote that kind of representational content, non-existent though it may be; and that term, to reiterate, will be 'non-propositional content.'

[86] In *Mind and World,* MacDowell famously argued that all representational content *can* be embodied in propositions.

[87] See Sellars [1].

[88] pp.26-32.

[89] Byrne [1] pp. 206-207.

[90] Byrne then formalizes the argument just given as follows. To ensure that I don't misrepresent Byrne's views, I will quote his own words:

Premise A

Assume the following: A subject enjoys two consecutive experiences: e, that ends at t, and then e. e and e* differ in phenomenal character. The subject is competent: if she tries to retrieve information (or misinformation) in memory, then her attempt will be successful, and she has no other cognitive shortcoming. The subject tries to detect a change in phenomenal character after t. The subject's memory is perfect: all the (mis)information supplied to her by e is available when she later enjoys e*.*

Then, solely on the basis of her current experience e and the memory produced by her past experience e, the subject will notice a change after t: if she gives an accurate report when e* is over, she will say that there was a change in phenomenal character after t. (pp.207-208)*

Premise B

Assume that a subject enjoys an experience e that ends at t and then experience e, and that after t the subject notices a change in phenomenal character, solely on the basis of her current experience e* and the (perfect) memory produced by her past experience e. Then the way things seem to the subject when she enjoys e differs from the way things seem when she enjoys e*. That is the content of e differs from the content of e*. (p. 210)*

Conclusion C

If a competent subject whose memory is perfect and who tries to detect a change in phenomenal character enjoys consecutive experiences e and e that differ in phenomenal character, then the content of e differs from the content of e**. (p. 213)

[91] I argue elsewhere that Byrne's argument doesn't even show this.

[92] See Peacocke [1] and [3].

[93] In fact, the only prominent author who does *not* hold this is Jerry Fodor; and Fodor does not consider the argument just given. See Fodor [1].

[94] To my knowledge, the position that representational capacities are non-atomistic – that they come only in articulated bundles – was first defended by Wittgenstein in *On Certainty*. (See Wittgenstein [1], especially section 141: "When we first begin to *believe* anything, what we believe is not a single proposition, it is a whole system of propositions. (Light dawns gradually over the whole.)" Wittgenstein's emphasis.) It – or a position similar – was famously given a brilliant defence by Wilfred Sellars. Some form of non-atomism is employed in *connectionist* theories of the mind. (See Bechtel [1].) These days, the prevalent view does seem to be that some form of non-atomism (holism or molecularism) is correct. But recently Jerry Fodor mounted a defence of atomism. (see Fodor [1]) Fodor's argument does not address the argument set forth here for representational non-atomism and, so far as I can tell, nothing Fodor says impugns this argument.

[95] Perhaps some anatomical corroboration for this thesis might be discovered. For example, perhaps the anatomical structures subserving a morphine induced high are in a different part of the brain from those subserving more paradigmatically representational cognitive faculties.

[96] See e.g. Byrne [1] p.227.

[97] In this and the following sections, the objections to my argument I will consider come from the mouths of hypothetical, not actual, interlocutors. The reason for this is that the argument I am currently running – namely, that intentionalism is committed to some kind of dubious representational atomism – has not yet appeared in the literature; so that there are, as of yet, no *actual* objections to that argument. Consequently, I must consider hypothetical objections to it.

[98] I should reiterate a clarification made earlier. In the literature, what I am calling 'propositional content" is typically referred to as "conceptual" content, and what I am calling "non-propositional" content is generally referred to as "non-conceptual content." (See e.g. MacDowell [1].} But, as I said in an earlier footnote, this nomenclature is deeply misleading. For it is generally agreed propositions can have things other than concepts as their constituents, and thus can have something that might legitimately be called "non-conceptual" content. Now if this is correct, then propositional content – i.e. the kind of content that can be embodied in propositions – is sometimes something other than a concept, and is thus sometimes non-conceptual. So the distinction between conceptual content and non-conceptual content is not the same as the distinction between propositional content and non-propositional content. Consequently, the terms "conceptual" content and "non-conceptual" content should not be used to mark the distinction between propositional content and non-propositional content.

[99] A number of philosophers hold that there is such a thing as non-propositional content. (See footnote 7.) Not all of those philosophers hold that phenomenal content supervenes on *any* kind of representational content (see footnote 7) and *a fortiori* do not hold that phenomenal content supervenes on non-propositional content. But *some* believers in non-propositional content hold that phenomenal

content supervenes on non-propositional content.; in particular, Tye is in this category. See Tye [1] and [2].

[100] In my view, only propositions can be thought: it isn't, strictly speaking, possible to think sub-propositional entities in isolation. But let us momentarily idealize away from that fact, if it is a fact.

[101] A classic discussion of the difference between real and simulated thought is found in Searle (1980). Searle's "Chinese room" experiment is meant to show that something can perfectly simulate the ability to speak Chinese without actually knowing Chinese. A man inside a room is given orders to make certain ink-marks when given other ink-marks. Unbeknownst to the man, the ink-marks in question are Chinese sentences, and the sequences of ink-marks form a meaningful and coherent discourse. According to Searle, the man's behavior *simulates* knowledge of Chinese without embodying any such knowledge. Therefore a simulation of knowledge doesn't add up to knowledge. The main response is to say that, while the man *per se* doesn't know Chinese, the system of which he is a part *does* know Chinese. (See Fodor 1980.) Searle counter-responds by describing *other* systems (e.g. robots) that simulate knowledge of Chinese but that, to Searle's mind, *clearly* don't have such knowledge. Searle's general approach is to say: (a) it is intuitively clear that the system described doesn't know Chinese, even though (b) that system simulates knowledge of Chinese; therefore (c) simulated knowledge (or thought) isn't actual knowledge. I happen to agree with Searle. But the problem with this line of thought is that the opponent can always generate contrary intuitions; he can say "I think that the system in question (e.g. the man with the translation manual, the robot) *does* know Chinese; I think you are begging the question against my viewpoint by saying, on the mere strength of your intuitions, that the system in question doesn't think: if the system reliably produces the right representations, then *ipso facto* it thinks." (See Fodor 1980.) What we need is some way to break this circle of bare intuitions. In this paper, I will try to provide such a way.

[102] In this paper, when I talk about "thought", I mean some mental entity, something spatio-temporal and presumably realized by neural events. I am not using the term "thought" to denote propositions or other (Fregean-Platonic) abstracta.

[103] It is, of course, generally held that representational content is causally effective. But, surprisingly, a number of otherwise seemingly reasonable positions appear to be inconsistent with this. For example, these days it is widely held that mental properties are "multiply realizable" (Putnam 1967, Fodor 1968, Davidson 1970, Lewis 1972). Causal role, not material composition, is said to be what realizes mental properties. Jaegwon Kim (1993: 339-367) has argued compellingly that this position renders mental properties causally inert – mental phenomena become *epi*phenomena. The underlying physical properties are doing the causal work. The mental phenomena don't *add* anything to the causally effective underlying processes. But, at the same time, they are not (ex hypothesis) *identical* with those processes. So, argues Kim, to avoid "epiphenomenalizing" the mental, we have to see mental properties not merely as *realized by*, but as *identical with*, physical phenomena. Kim rejects the idea that the mental could be "de-epiphenomenalized" by positing "downward causation". If the mental is realized by something non-mental, then the mental is powerless. So, in Kim's view, given that the mental is not powerless, the relation between it and brain-events is one of identity, not supervenience or realization. In any case, Kim has exposed a genuine *prima facie* tension between, on the one hand, the presumption that mental properties are effective and, on the other hand, the truth of the multiple-realizability thesis.

Kim comes very close to saying that, unless mental entities have causal powers, they don't exist at all! He says "to be real is to have causal powers" (Kim 1993: 348). So any kind of epiphenomenalism – and, in Kim's view, that means anything other than "reductive" materialism – is inconsistent with the very *existence* of mental entities.

Kim's dictum (to be real is to have causal powers) appears to have a far-reaching consequence. If representational content is *real*, then it has causal powers; and it doesn't have such powers, it isn't real. Some forms of externalism would appear to render representational content causally inert, and it would

be interesting to see how they deal with Kim's suggestion that spatiotemporal reality involves causal efficacy. McGinn (1998: Chapter 2) makes a concerted effort to deal with this very problem.

Frank Jackson and Philip Pettit (2004) discuss a related issue that arises in connection with functionalism. According to functionalism b (an arbitrary brain-state) has such and such content wholly in virtue of its causal role. On the face of it, functionalism seems to *give* causal properties to the mental: it makes causal properties essential. But, at least arguably, the exact opposite is the case. Suppose that *b* encodes the proposition *snow is white.* If b encodes that belief *in virtue of its causal role*, then its encoding that belief cannot be what *causes* b to behave as it does. If content supervenes on causal role, then causal role is more basic than, and is independent of, content. So, ironically, functionalism would seem to *strip* the representational of causal efficacy. (This has a certain similarity to Kim's point.) Jackson and Pettit argue that, notwithstanding this point, functionalism *is* consistent with the potency of representational content. The relevant point here is that functionalism – a reasonable, some would say an *inevitable*, view – is, at least arguably, in tension with the causal effectiveness of the mental; and that tension becomes apparent only when we distinguish *b* itself from *b's* having such and such representational content.

Many have thought it strange to deny that the representational is causally effective (for example, see McGinn 1998: Chapter 2). But nobody has said (though Kim came close: see above) that representational content must effective if there are to even *be* thought-processes or therefore (since a single thought is presumably a limiting case of a thought-process) any thought at all. That is part of what I will argue here. I will argue that, unless representational content is effective, representations cannot "molecularize" into thought-processes, or therefore into single thoughts.

[104] Davidson 1973: 79.

[105] Gettier 1963.

[106] Here I am using these terms in a broad sense. I am using the term "logical" to cover not just *formally* but also *informally* correct reasoning.

[107] Before going any further, let us be clear on one thing: there *will* be such a property as p. It does seem pretty clear that, in at least some cases, our brain-states have representational content. But not *every* physical property of a brain-state will be involved in its mediating a certain representational content. Among other things, it would make possession of representational content far too fragile. Let b be a brain-state (or, better, an enduring brain-condition) that mediates your standing belief that two is even. Like any other spatiotemporal structure, b has many different causal properties. And like practically any other spatiotemporal structure, b will, in the course of its existence, undergo many changes in those properties. It will undergo changes in its specific shape, its specific mass, its specific location, its specific temperature. All of these are causally effective properties. If b should have temperature 99.8674345 as opposed to 99.8674346, that is a causally significant fact: it will change, however minutely, what b does to certain other things.

Clearly b can undergo changes – perhaps not overly drastic changes, but some changes – in these properties *without* ceasing to mediate a belief that two is even. In the time you've been reading this sentence, every one of your brain-states has undergone innumerable, causally significant changes: no one of those brain-states has kept *exactly* the same temperature throughout that interval; no one of them has kept *exactly* the same micro-morphology. But most, if not all, of them have continued to do their *representational* jobs. You still believe that Paris in France; you still believe that snow is (usually, more or less) white. The brain-structures that realized these beliefs two seconds ago are numerically identical with the ones that now realize those beliefs, even though, as we just said, those brain-states have undergone innumerable (small, but non-Cambridge) changes.

So *some* of B_1's causally effective properties can be changed without jeopardizing B_1's ability to

mediate a belief that grass is green. It follows that not *every* one of B_1's causally effective properties is involved in B_1's realizing a belief that grass is green. We have thus established that p exists, where p is some causally effective property of B_1 that is not involved in B_1's mediating a belief that grass is green.

[108] Nonetheless, even if your thinking *Socrates* is what causes you to think *hit* and *Plato* (in that order), it *still* doesn't follow that you have thought *Socrates hit Plato.* We could imagine an analogue of the scenario discussed earlier. For some reason, whenever you have a Socrates-thought, your heart starts to race (because, let us suppose, you love his works so much). A hypnotist had previously programmed you to think *hit* and *Plato* (in that order) whenever your heart-rate goes above 90 beats per minute. So you have b_1, and b_1's being a Socrates-thought causes you (however "deviantly") to think *hit* and then *Plato.* Still, even under that circumstances, you have not thought *Socrates hit Plato.* You have thought *Socrates* and then *hit*, and so on.

[109] I will refer to this position as CTM* in order to mark its obvious similarity to the computational theory of mind. But the "*" is meant to show indicate that I wish to leave it open, at least in those context, whether CTM* is quite the same thing as the computational theory of mind.

[110] See the Introduction to Fodor 1981a, and also Fodor 1981b, 1994a, 1994b, and 1987 (Chapter 1, especially pages 18-20).

[111] Fodor (1990a: 22) "You connect the causal properties of a symbol with its semantic properties via its syntax…To a first approximation, we can think of its syntactic structure as an abstract feature of its (geometrical or acoustic) *shape* [Fodor's emphasis]. Because, to all intents and purposes, *syntax reduces to shape* [my emphasis], and because the shape of a symbol is a potential determinant of its causal role, it is fairly easy to see how there could be environments in which the causal role of a symbol correlates with its syntax. It's easy, that is to say, to imagine symbol tokens interacting causally *in virtue of* [Fodor's emphasis] their syntactic structures. The syntax of a symbol might determine the causes and effects of its tokenings in much the way that the geometry of a key determines which locks it will open."

[112] This idea really goes back to Russell (1905/1990a). Richard Montague used it to great advantage in his work on natural language quantification. See MacCawley (1981a: Chapter 14) for a brief, clear discussion of some of the relevant aspects of Montague's work.

[113] At least as it occurs in (1). It may have different meanings in different contexts.

[114] See Fodor 1994a (22-23):

"[W]e know from formal logic that certain of the semantic relations among symbols can be, as it were, "mimicked" by their syntactic relations; that, when seen from a very great distance, is what proof-theory is all about. So, within certain famous limits, the semantic relation that holds between two symbols when the proposition expressed by the one is implied by the proposition expressed by the other can be mimicked by syntactic relation in virtue of which one of the symbols is derivable from the other. We can therefore build machines which have, again within famous limits, the following property: the operations of such a machine consist entirely of transformations of symbols; in the course of performing these operations the machine is sensitive only to syntactic properties of the symbols; and the operations that the machine performs on the symbols are confined entirely to alterations of their

shapes. Yet the machine is so devised that it will transform one symbol into another if and only if the symbols so transformed stand in certain *semantic* relations; e.g. the relation that the premises bear to the conclusion in a valid argument. Such machines – computers, of course, just *are* [Fodor's emphasis] environments in which the causal role of a symbol is made to parallel the inferential role of the proposition it expresses."

Fodor hedges his points with phrases like: "within certain famous limits", "as it were", "from a very great distance", and "again within famous limits". But, as I hope to show, these hedges turn out either to be empty or to conceal systemic problems with Fodor's analysis.

[115] For the record, this is not my own view. I think that the syntactic structure of a sentence is so deeply in facts about its semantics, that it is *not* clear whether "Bill is a green duck, therefore Bill is duck" has the same syntax as "Bill is a decoy duck, therefore Bill is a duck". Chomsky (1965) argues persuasively that "John expected the visitors to leave" has a very different syntax from "John wanted the visitors to leave". Obviously this point is meant to generalize. If it does, then we must rethink the traditional view that, in English (and other natural languages), there is a misalignment of syntax with the semantics.

[116] See Davidson (2004).

[117] See Putnam (1975). The contents of the next few paragraphs are a very watered down version of a famous thought-experiment that Putnam gives in that publication.

[118] Putnam talks more about our concepts of natural kinds than about our concepts of individuals. But in the present context, what is important is that one's concept of spatiotemporal individuals *or* kinds has a causal component; and this point is due, I believe, to Putnam. In the present context, the distinction between our concepts of individuals and our concepts of kinds is not particularly important.

[119] Where the objects are abstract, the causal connection is between ourselves and *instances* of those objects. So my concept of the number two consists in a certain kind of causal connection between brain-states of mine and instances of the number – e.g. pairs of shoes or pairs of hands.

[120] Though Davidson may not go so as to state explicitly that this is his view.

[121] Of course, when I say that M responds to inputs 'in exactly the way' that a thinking being responds to those inputs, I do not mean that M necessarily has the same *thoughts*; I mean that his behavioral output – the souds and movements he produces – are those of a thinking being. For if I stipulated that M's *cognitive* reactions were just like those of a normal human being, then Turing's question – does M think? – would be *trivial*; for it would be *built into* the way we described the situation that M thinks.

[122] This, in effect, is the tack Davidson takes in another publication.

[123] In this paper, when I talk about 'brains in vats', I mean brains that are just like the brains of *living*, cognitively normal people, and are subject to the some proximal, but not the same distal, stimulations. So the brains in vats I will be talking about are just like my brain and your brain – except that they are in vats. Of course, in reality, there are brains in vats. But they are dead. They are not what I mean by 'brains in vats.'

[124] Of course, your dream might symbolically correspond to something real: the unicorn might be your mother, the griffin might be your father. I am talking about manifest, not latent, dream-content. Also, one can obviously dream about real things (e.g. one's father). But one *needn't* dream about such

things. I am discussing dreams that are wholly about unreals like goblins.

[125] When I talk about 'object-dependent propositions', I mean propositions whose existence is dependent on the existence of *spatiotemporal* individuals and kinds, e.g. *Socrates is bald*, *water quenches thirst.* I *don't* mean propositions whose existence depends on that of *platonic* individuals and kinds (e.g. the number two). So *2+2=4* is not 'object-dependent', as I am using this expression, even though (perhaps) its existence is dependent on that of certain objects, namely *two* and *four* and so on. My usage of the term 'object-dependent' is quite conventional.

[126] In any case, this is that Evans (1984) persuasively argues for.

[127] Colin McGinn (1988: 50-51) ably defends a very similar point. McGinn talks about artifact concepts (e.g. *table*). But what he says in that connection applies, it seems to me, to functionally defined concepts like *leg*.

[128] Fodor (1981a: 2-3).

[129] This is arguably subject to some qualifications associated with externalism. But these will not come into play in the present paper.

[130] Fodor (1981a: 11-16); Fodor (1994: 294); Fodor (1987: 18-19).

[131] Once again, this is arguably subject to some qualifications relating to epistemic externalism that won't play a significant part in the present discussion.

[132] One *conceivable* way out is to say that our mental processes are over-determined: they are determined *entirely* by the laws of physics (for they are realized by brain-events, and brain-events are determined by the laws of physics). But they are *also* determined by consciousness of norms.

But this is a radically implausible thesis. Let E be an arbitrary brain-event that is identical with, or that realizes, a mental-event. For the sake of argument, suppose that E is "over-determined" in the relevant sense. This would mean that E was *entirely* caused by some state of affairs S (some purely physical or physiological state of affairs) and *also* that E was caused by some state of affairs S* (some awareness of the canons of rationality), such that S and S* were completely distinct. But the idea of an event's being *entirely* caused by two distinct states of affairs is of doubtful coherence. Also, if *each* brain-event that realized a mental event were *entirely* caused *both* by some awareness of norms and *also* by some physiological state of affairs, then we'd have innumerable cases of two entirely different kinds of things eventuating in the exact *same* thing: we'd have innumerable cases where two entirely distinct things – one a consciousness of norms, the other a physiological state of affairs – independently eventuated in the exact *same* thing (some brain-event). But to accept this is to accept innumerable unexplained coincidences. By contrast, if we suppose that mental states of affairs *are* neuro-cerebral states of affairs, then it is no longer a coincidence that brain-events realizing mental entities have physiological causes *and* mental causes: the former *are* the latter, at least in some causes. So it would not be good methodology to accept the idea that brain-states are over-determined. And practically no one does accept. (I myself do not.)

[133] I should make one point very clear. I am not opposed to the idea that *some* thought consists in computation. I could perform a computation right now. And I am not opposed to the idea that many of our cognitive achievement derive from sub-personal acts of computation (like those posited by Chomsky and Marr). What I am opposed to is the idea that computation might constitute the *foundation* of cognitive activity. Computation, I will argue, exists only in a rich background of cognitive and representational accomplishment. So even though computation might be a pervasive feature of our cognitive lives, it cannot be the *basis* of it; it necessarily has a sub-ordinate (though perhaps deeply important) role in cognition.

[134] Fodor 1987 (chapter 1, especially 18-20); Fodor 1981 (*introduction*, especially 22); Fodor 1990a; Pylyshin 1984.

[135] Fodor 1994a (86); Fodor 1987 (chapter 1). See also Fodor 1981 (22): "[M]achines operate on symbols in virtue of their formal (not their semantic) properties…" If the term "formal" means "morphological", then form-driven operations are not cognitive at all. If the term "formal" means "syntax-driven", then "formal but not semantic" is simply a contradiction in terms, as we will see.

[136] Fodor 1981a: 226. "[C]omputational processes are both *symbolic* and *formal*. They are symbolic because they are defined over representations, and they are formal because they apply to representations in virtue of (roughly) the *syntax* of the representations." A few paragraphs later (227), Fodor writes: "[F]ormal operations apply in terms of the, as it were, shapes of the objects in their domain." This passage contains the hedge "as it were". But on any interpretation of this hedge, as we will see, there simply doesn't exist the kind of connection between syntax and morphology (shape) that Fodor's position alleges.

[137] See Fodor (1981a: Introduction); Fodor (1987: chapter 2); Fodor (1990a); Fodor 2001 (3-45).

[138] See Fodor (1987: 19) "The operations of the machine [on which I am modeling the human mind] consist entirely of transformations on symbols; in the course of performing these operations, the machine is sensitive solely to syntactic properties of the symbols; and the operations that the machine performs on the symbols are entirely confined to altering their shapes."

See also: Fodor (1990a: 22) "You connect the causal properties of a symbol with its semantic properties via its syntax…To a first approximation, we can think of its syntactic structure as an abstract feature of its (geometrical or acoustic) *shape* [Fodor's emphasis]. Because, to all intents and purposes, *syntax reduces to shape* [my emphasis], and because the shape of a symbol is a potential determinant of its causal role, it is fairly easy to see how there could be environments in which the causal role of a symbol correlates with its syntax. It's easy, that is to say, to imagine symbol tokens interacting causally *in virtue of* [Fodor's emphasis] their syntactic structures. The syntax of a symbol might determine the causes and effects of its tokenings in much the way that the geometry of a key determines which locks it will open.

But, now, we know from modern logic that certain of the semantic relations among symbols can be, as it were, 'mimicked' by their syntactic relations; that, when seen from a very great distance, is what proof-theory is about. So, within certain famous limits, the semantic relation that holds between two symbols when the proposition expressed by the one is entailed by the proposition expressed by the other can be mimicked by syntactic relations in virtue of which one of the symbols is derivable from the other. We can therefore build machines which have, again within famous limits, the following property:

The operations of the machine consist entirely in transformations of symbols;
in the course of performing these operations, the machine is sensitive solely to syntactic properties of the symbols;
and the operations that the machine performs on the symbols are entirely confined to altering their shapes.

Yet the machine is so devised that it will transform one symbol into another if and only if the propositions expressed by the symbols that are so transformed stand in certain semantic relations – e.g. the relation that the premises bear to the conclusion in a valid argument. Such machines – computers, of course, -- just *are* [Fodor's emphasis] environments in which the syntax of a symbol determines its

causal role in a way that respects its content. This is, I think, a pretty terrific idea; not least because it works."

Contrary to what Fodor holds, we will see that building a syntax-driven machine is difficult as building a semantics-driven machine. For syntax is an aspect of semantics; and the connection between the morphology of a symbol and its syntax is as remote and conditional as the connection between a symbol's morphology and its semantics.
See also: Fodor (1981a: 13-17); Fodor (1994: 294-298); Fodor (1987: 18-20); Fodor 1990a; and Pylyshin (1984).

[139] Expression-tokens can be spoken or realized by hand-signals or light-beams. They needn't be realized by ink-marks or by anything else that has "shape" in the narrow sense. In this paper, I will use the term "morphology" in a generalized sense – to describe acoustical (or hand-signal) *counter-parts* of the physical shape had by written expression-tokens.

[140] Compare Fodor (1981: 226-227). Having written that "[mental processes] are formal because they apply to representations in virtue of (roughly) the *syntax* [Fodor's emphasis] of the representations" – in other words, having identified "formal" with "syntactic" operations -- Fodor then goes onto write: "Formal operations are the ones that are specified without reference to such semantic properties of representations as, for example, truth, reference, and meaning."
This, I will argue, is dead wrong. Formal, in the sense of syntactic, operations are a kind of semantic operation.

[141] I should reiterate that this is in no way meant to conflict with the important discoveries of Chomsky and Marr. Obviously people *can* compute; and people may well engage in the sub-conscious (and sub-personal) computations posited by Chomsky and Marr. So in *that* sense, computations may have a very important role in our cognitive lives. What I am saying is that, if Chomsky and Marr are right (as they may well be), the computations they posit presuppose an even more fundamental stratum of cognitive functioning: for any computation, in a relevantly robust sense of the word, presupposes a considerable cognitive and representational background, as I will argue.

[142] See Kaplan 1977, 1989; Barwise and Perry 1999.

[143] Quine (1970: 58) wrote: "[A] logical truth is a truth that cannot be turned false by substituting lexicon [lexical items] for lexicon." In other words, a logical truth is one that cannot be made false by replacing *lexical items* with *lexical items.* And a "lexical item" is defined as one that *picks something out* or, at any rate, is *supposed* to pick something out, given its grammatical role. ("Socrates" and "fat" are lexical, because they pick out an individual and a property, respectively. "Pegasus" is lexical because it is *supposed* to pick something out; its grammatical role is like that of "Socrates".) Carnap (1947: 85-88) described any sign that was "descriptive" – i.e. that picks something out or expresses a property – as being "non-logical". In his view, purely "logical" terms – by which he means 'or', 'not', 'and' and so on – do not pick out or describe or express anything. They are purely "formal". We will take issue with this view in due course.

[144] These ideas are found, in some form, in Russell (1905). (Russell himself denies that quantifiers refer. But much of the contemporary inspiration for the view that they refer to second-level functions is due to Russell 1905.) Ultimately, these ideas go back to Frege (1892a, 1892b). Richard Montague (1974) used them to great advantage in his work on natural language quantification. See MacCawley (1981a: Chapter 14) for a brief, clear discussion of some of the relevant aspects of Montague's work.

[145] Frege 1892a, 1892b.

[146] Montague 1974, MacCawley 1983.

[147] There is *a* sense in which a spoken of token of "I am tired" is isomorphic with a written token and with a token in hand-signals. But *that* sense of "isomorphic" does not have to do with *morphology*; it has to do with semantics. *Given* that any two of those three tokens have certain *semantic* similarities, the written "I" can be put into a certain kind of correspondence with its spoken (or hand-signaled) counterpart, the same being true (mutatis mutandis) for the spoken "tired", the spoken "am", and so on. But *absent* that semantic similarity – if we just focus on physical morphology – those three tokens have about as little in common as any three things. Any "isomorphism" that holds between them is underwritten by semantic similarities, not (mere) morphological similarities.

[148] See Kaplan 1978, Kaplan 1989 (586-587), Barwise and Perry 1999 (32-44).

[149] See Barwise and Perry 1999 (43-44). They make exactly this point.

[150] This point is clearly made by Barwise and Perry 1999 (32-44). It is also made by Kaplan (1989). It ultimately goes back to Strawson (1951).

[151] Barwise and Perry (1999: 32-39) deserve credit for making this point very clearly; they used the term "efficiency" to mark the feature of language which I am discussing.

[152] See Kaplan 1989 (557-558) and Barwise and Perry 1999 (32-34).

[153] See Kaplan 1989 (557-558).

[154] See Kaplan 1989 (557-558).

[155] See Barwise and Perry 1999 (32-34).

[156] My own view is that, even here, there is no purely syntactic characterization of logical truth. It is true that, in (a)-(c), the indexicals are not important. But the *reason* they are not important has to do with the peculiarities of the concepts involved – it has to do with the fact that, whether members of a species are mortal is not a function of time: if members of a species are mortal at t, they will be mortal at time t*. That fact, in its turn, turns on peculiarities of the property of mortality. Not all properties are like that. For example, whether all members of a species are thirsty or are in Antarctica *is* a function of time. The relevant differential facts about the concept of mortality are not coded in the *syntax* of (a)-(c). So the irrelevance of the tense-markers in (a)-(c) is a function of subtleties buried in the *semantics* of the terms involved in that argument. So even where artifacts like (a)-(c) are concerned, there is a gap between syntax and logical form, and that gap is bridged only by semantics.

[157] Of course, this term is borrowed from Fodor 1975.

[158] See Barwise and Perry 1999 (43-45).

[159] See Barwise and Perry 1999 (27-44).

[160] See Barwise and Perry 1999 (45).

[161] Conceivably, one might take a line like the following:

"Consider the sentence-token:

(*)"Bob now looks tired, and Bob now does not look tired".

The syntactic form of (*) is a function of whether the two occurrences of "Bob" co-refer; and it is a function of whether the indexicals ("now", "-s") refer to the same bit of time. If the expression in question *don't* co-refer, (*) has one syntactic form; if they *do* co-refer, it has some other syntactic form. So the facts you mention about indexicality, and about ambiguity, do *not* create a gap between syntactic and logical form. They show only that syntactic form is more fine-grained than you allow."

First of all, this response is completely ad hoc, and is contradicted by everything we know about the concept of syntax. But even if correct, this response isn't available to CTM; for it amounts to saying that syntax is even *more* a matter of semantics than even we previously said. The objector is saying that the very *syntax* of (*) is a function of the specific referents of the indexicals, and of the singular terms, in it. So, if the objector is right, to know the syntactic form of (*), it is not enough to know generic facts about the semantics of its expressions; it is not enough to know that "Bob" is a singular term, and that each occurrence of "now" refers to time of utterance: one has to know to *which* specific person "Bob" refers, and what the exact times of utterance are. The objector's response thus makes syntax be totally inaccessible to anything that lacks fairly high-resolution knowledge of semantics; it thus makes syntax-drivenness *presuppose* the existence of various representational and cognitive abilities. Thus the objector's point makes it even harder for syntax driven operations to form the foundation of our cognitive lives.

[162] See Fodor 1994a (22-23):

"[W]e know from formal logic that certain of the semantic relations among symbols can be, as it were, "mimicked" by their syntactic relations; that, when seen from a very great distance, is what proof-theory is all about. So, within certain famous limits, the semantic relation that holds between two symbols when the proposition expressed by the one is implied by the proposition expressed by the other can be mimicked by syntactic relation in virtue of which one of the symbols is derivable from the other. We can therefore build machines which have, again within famous limits, the following property: the operations of such a machine consist entirely of transformations of symbols; in the course of performing these operations the machine is sensitive only to syntactic properties of the symbols; and the operations that the machine performs on the symbols are confined entirely to alterations of their shapes. Yet the machine is so devised that it will transform one symbol into another if and only if the symbols so transformed stand in certain *semantic* relations; e.g. the relation that the premises bear to the conclusion in a valid argument. Such machines – computers, of course, just *are* [Fodor's emphasis] environments in which the causal role of a symbol is made to parallel the inferential role of the proposition it expresses."

Fodor hedges his points with phrases like: "within certain famous limits", "as it were", "from a very great distance", and "again within famous limits". These hedges turn out either to be empty or to conceal systemic problems with Fodor's analysis.

The "famous limits" Fodor has in mind have to do with incompleteness results, the most famous being the two of Gödel's. An apparent (and, in the view of most, an actual) implication of these results is: the bearing relations that hold among number-theoretic statements cannot be reconstituted within the confines of the concepts corresponding to so-called syntax-defining terms like "not", "or", "and", "for some", and the like. Admittedly, the implications of such incompleteness theorems is a matter of great controversy.

But such exotic results in mathematical logic are the *least* of the problems affecting a mirroring of semantics in syntax. Remember what we said earlier. The syntax of a sentence is a function of the more generic facts about its semantics. "Bob is tired" has a certain syntax because "Bob" is a singular term, "tired" is a predicate (a one-place function), and so on. For this reason "Tim is angry" and "Bob is tired" have the same syntax. Logicians have created (or outlined) languages in which, supposedly, logical form is made perspicuous by syntactical structure. But no matter how perspicuous a language is (from the view point of its syntax reflecting its semantics), sentences having very different semantics will have the same syntax. In any language having any expressive power at all, there will be semantically different sentences having the same syntax. So a mind that tracks syntax will inevitably fail to do justice to semantic relations.

(Henceforth, to facilitate discussion, sentences in **boldface** are supposed to be the Mentalese translations of their English counterparts.)

Fodor would not necessarily be moved by these points. For he seems to have held that "**Mary is nervous**" and "**Fred is angry**" *do* have different syntaxes. His reason, if I understand him correctly, is they have different *morphologies* and morphology is, or reduces to, syntax. (This point is crucial to Fodor's way of dealing with Frege-cases and also with the apparent truth of externalist theories of content, like Putnam's. See Cain 2002 (135); Fodor 1994d.)

But, as we've seen, morphology doesn't have *remotely* the kind of connection to syntax that it would have to have for "**Mary is nervous**" and "**Fred is angry**" to have different syntaxes. And, plainly, they have the same syntax, the reason being that they are like in the relevant semantic respects ("**Mary**" and "**Fred**" are both singular terms, "**angry**" and "**nervous**" are both one-place predicates, and so on). Perhaps Fodor can circumvent this problem. But it is quite clear that, for any one syntactic form, there are many sentences having that form that have different semantics. And this seriously problematizes the idea, so crucial to CTM, that semantics can be mirrored in syntax.

In response, one could say (in effect, Fodor (1994d) *does* say) that the morphological differences between "**Mary is nervous**" and "**Fred is angry**" give them different causal roles, explaining why they lead to different beliefs.

But this leads back to the problem we were discussing earlier (pages 2-3). A morphology-driven operation is not cognitive at all – is merely *simulated* thought -- except in so far as syntax (or semantics) is realized by morphology. But morphology, by itself, to no degree realizes syntax (or semantics), as we've seen time and time again. So if the differences in causal role between "**Mary is nervous**" and "**Fred is angry**" are purely morphology-based, then those differences are not *cognitive* at all. Quite plainly, there is a *cognitive* basis to my willingness to infer *somebody is nervous or tired* from *Mary is nervous*, but not from *Fred is angry*. So if, as Fodor supposes, our thoughts are mediated by Mentalese expressions, then there is a *cognitive* basis to the different causal roles had by "**Mary is nervous**" and "**Fred is angry**". So the difference between "**Mary is nervous**" and "**Fred is angry**" *cannot* be purely morphology-based.

Conceivably, Fodor counter-reply by saying this:

"For the reasons you give, where symbols of natural language are concerned, any shape can be associated with any syntax at all, given the right background facts. But things are different where Mentalese symbols are concerned. In Mentalese, morphology *does* realize syntax."

Since, as we have seen, syntax is an aspect of semantics, this reply amounts to saying: "In Mentalese, morphology realizes semantics."

But no advocate of CTM, and definitely not Fodor, would say that morphology realizes *semantics*, even where Mentalese is concerned. CTM is a form of functionalism. Functionalism says that the semantics of a brain-state supervenes *not* on its morphology but on its causal role. So given that syntax supervenes on semantics, functionalism says, in effect, that the *syntax* of a brain-state does not supervene on its morphology.

(In many places, Fodor vehemently argues that the semantics of brain-states don't supervene on *any* of their intrinsic properties, let alone on their shapes. So Fodor himself would never say that, where Mentalese is concerned, semantics supervenes on morphology. See Fodor 1981b, 1994a-1994d, 1998.)

[163] Fodor 1994a (23) writes:

"There must be mental symbols because, in a nutshell, only symbols have syntax, and our best available theory of mental processes…needs the picture of the mind as a syntax-driven machine."

Fodor is quite right to say that only linguistic expressions have syntax. So he is quite right to say that *if* thinking consists in syntax-driven operations, *then* there is an internal language (a "language of thought"). But, unfortunately for Fodor's position, syntax doesn't adequately reflect logical form, for the reasons we've just been considering. So if we were syntax-driven machines, we would be logic-blind.

[164] Barwise and Perry 1999 (22).

[165] This point, or one like it, is famously found in Wittgenstein 1958, though Wittgenstein's point seems to be less mentalistic. Similar points are found in Putnam 1975 and Burge 1979.

[166] A line of thought concerning linguistic meaning that has certain similarities to this one – but that has a far less mentalistic orientation – is found in Wittgenstein 1958.

[167] Since Frege, it has been generally assumed that, in natural language, syntax is a poor indicator of logical form. Consider:

(a) "Bill is a green duck, therefore Bill is a duck."
(b) "Bill is a decoy duck, therefore Bill is a duck."

The stock line on this sort of phenomenon is as follows. In (a), syntax tracks logic; in (b), syntax does not track logic. But (a) and (b) have the same syntax. Therefore syntax does not track logic, at least not in natural language. (In artificial languages, syntax may reflect logical form.)

By itself, this is not a problem for CTM. CTM can say:

Thought consist in syntax-driven operations; but those operations involves symbols of what we might

call a "well-behaved" language – the kind that Frege and Russell talked about: a language in which there is a kind of coincidence between logical and syntactical form. In English "Bill is a decoy duck" has the same syntax as "Bill is a green duck." But "Bill is a green duck" means: "Bill is a duck and Bill is green", whereas "Bill is a decoy duck" does not mean "Bill is a decoy and Bill is a duck". If thinking consists in syntax-driven operations, the medium of computation is a language in which doesn't contain distortion like "Bill is a decoy duck" – everything is made perspicuous.

This is a reasonable response; but it leaves some important issues untouched. No speaker of English would regard (b) as a legitimate inference. Why not? Because anyone who knows the meaning of (b) knows that it is not equivalent to: "Bill is a duck and Bill is a decoy, therefore Bill is a duck". A knowledge of *meaning* is what prevents faux-inferences like (b).

But a knowledge of meaning is *always* involved in determinations of syntactic structure, even in cases of the idealized languages referenced by Russell and Frege. Suppose we spoke a language like English except that, in this other language, syntactical form mirrored logical form. Let English* be this other language. The English*-translation of "Bill is a green duck" is "Bill is green and Bill is a duck". The English*-translation of "Bill is a decoy duck" is *not* "Bill is a decoy and Bill is a duck"; it is (roughly) "Bill is an object that looks like a duck, but is not a duck". So English* doesn't contain the devices that, in English, make syntax diverge from logical form.

We know that, in *actual* English, a knowledge of semantics is needed to avoid faux-inferences like (b): if one relies only on syntax, one is thrown off. But the same is *always* true; it is true of English* no less than English. Consider the sentence:

(c) "Bill is green and Bill is a duck, therefore Bill is a duck."

This sentence is valid; and its validity is (reasonably well) reflected in its syntax. (I say "reasonably well" because, for reasons discussed earlier, the context-sensitive components in the sentence prevent its logical form from being completely reflected in its syntax.) But you couldn't possibly know what syntax it had unless you knew that "Bill" were a singular term (i.e. denoted an individual), "green" were a one-place predicate (i.e. denoted a function of one variable), and so on. If these terms had different denotations, then (c) would have an entirely different syntactic form. There is no knowledge of the syntax of (c) without knowledge of some facts about its semantics'. It doesn't matter whether (c) is considered as a sentence of English or English*. Taken as a sentence of English*, (c) is well-formed only if "Bill" denotes an individual, "green" is a one-place predicate, and so on. So even in English*, syntax-driven operations – so-called "formal" inferences -- are semantics-driven. The difference between English and English* is not that, in the latter one needs *no* semantic knowledge to make formal inferences, whereas, in the former, one does need such knowledge. The difference lies in how much semantic knowledge is needed. In English, one needs knowledge of the specific meanings of predicates to make the right inferences. In English*, one needs only fairly generic knowledge; one doesn't need to know the *specific* meanings of expressions; one need only know that (e.g.) "small" denotes *some* one-place predicate or other. But in any language, no matter how well-behaved, syntax-driven operations are semantics driven. When you say that a language is well-behaved, you aren't saying that, in it, the set of semantic facts is reflected in *non*-semantic facts; you are saying that the set of semantic facts is reflected in a certain subset of that same set.

There is another way to look at all this: a way that, if correct, is even more problematic for CTM. Again consider:

(a) "Bill is a green duck, therefore Bill is a duck."
(b) "Bill is a decoy duck, therefore Bill is a duck."

As we noted, the standard diagnosis is that, in English and other natural languages, syntactic

categories don't appropriately line up with semantic categories; and, for that reason, syntax-based inferences are often erroneous, when we are dealing with sentences of natural language.

This is certainly one way to look at things. There is another possibility. It is a distinct possibility that "decoy duck" and "green duck" *are* syntactically different. As we saw, the syntactic profile of a sentence is a function of the more generic facts about the semantics of its constituents: it is a function, not of what *specific* things its constituents mean, but of the *kinds* of thing they mean. "Jerry snores and Mary does not snore" has a certain syntactic structure; it has it because "Jerry" is a singular term, "snores" is a first-order one-place predicate, "and" is a certain kind of functor, and so forth. Given this, consider:

(A) "Something is tall".
(B) "Bob is tall".

"Bob" denotes an individual; it is a singular term. "Something" does not denote an individual. It is not a singular term. It is a second-level predicate, one that (arguably) denotes a second-level function. "Tall" (if Frege is right) denotes a function from individuals to truth-values: so it denotes a first order function. "Something" denotes a function from first order functions (functions from individuals to truth-values) to truth-values. Supposing that the Frege-Montague analysis is even remotely right, "something" denotes a function F that assigns truth to a first-order function f exactly if, for some x, f assigns truth to x. "Something" and "Bob" denote such different things that they are plausibly seen as falling into different syntactic categories; so (A) and (B) are plausibly seen as being syntactically different. Of course, (A) and (B) are *grammatically* parallel. But, as Chomsky pointed out long ago, grammatical categories make for very superficial taxonomy. Grammar is folk-syntax; grammatical descriptions are shallow and, at best, approximate, syntactic descriptions. (When I talk about "grammar", I am referring to the taxonomic system that we are taught in school: the system that includes categories like "noun", "adjective", "determiner", "conjunction", and that regards "a man" and "Socrates" as both being "noun phrases" and thus in the same syntactic category. So when I talk about "grammar", I am not using the term that it bears in expressions like "generative grammar", "universal grammar", or in any of the theories that aspire to replace the taxonomy we learn in school with a deeper characterization of language.)

It is a possibility worth considering that "decoy" and "green", like "Bob" and "something", denote entirely different kinds of things. If this is right, then

(C) "Bill is a green duck",

and

(D) "Bill is a decoy duck",

do not have the same syntactic structure. The sentence "Bill is a green duck" is true exactly if Bill is both green and duck. For (C) to be true, it is necessary that Bill fall into a sub-class of the class of ducks. So the term "green" can be seen as a function that maps a predicate (in this case "duck") onto a restriction of itself. But the term "decoy" does not function in this way at all. For (D) to be true, Bill must fall *outside* the class of ducks. So "decoy" does not denote a function from predicates onto restrictions of themselves. It has for its meaning an entirely different sort of function.

Given that "decoy" and "duck" denote such different things, it is not out of the question that they might be syntactically very different. Remember that the syntax of an expression consists in very generic facts about its semantics. Where "decoy" and "green" are concerned, those generic facts might be sufficiently different to warrant putting those expressions in different syntactic categories. If that is right – I am not saying that it is, but I am saying that it is a possibility worth looking into -- then the

syntax of (C) *is* entirely different from the syntax of (D). In that case, the difference between

(a) “Bill is a green duck, therefore Bill is a duck.”
(b) “Bill is a decoy duck, therefore Bill is a duck.”

is reflected in a difference in syntax. So the fact that (a) is valid, while (b) is not, *does not* necessarily reflect a failure for English semantic categories to line up with English syntactic categories. We are quick to see English, and other natural languages, as badly behaved. We take the invalidity of inferences like (b), and like “Bill is a small elephant, therefore Bill is small”, to indicate that English lacks a certain transparency – that it is, in some respect, “badly behaved”. What I am suggesting is that, arguably (though not definitely), those cases of bad behavior are merely apparent. When we look at the syntax of the expressions involved, we find that English is no less well-behaved than English* or any formal language. Syntactic facts are so deeply embedded in semantics facts that there no hope of a syntax-driven entity that isn’t also semantics-driven. I am suggesting that, in English, semantic and syntactic categories quite possibly do line up as well as they do in formal languages. The view that they do not is based on a confusion of grammar (in the sense of what we learn in middle school, not in the sense of “generative grammar”, “transformational grammar” and the like), which is a superficial theory about syntax, with syntax itself. Because of sentences like (b), it is obvious that there is no way for a semantics-blind entity to track logical relations among English sentences. So if English is well-behaved, as I am suggesting it might be, then it is obvious that a semantic-blind entity cannot track logical relations holding among sentences of a well-behaved language. Semantic knowledge will always be involved.

A different example might clarify my point. Traditional grammar sees the following two sentences as having the same form:

(A) “John touches a boat.”
(B) “John wants a boat”.

But there is every reason to believe that “wants” and “touches” are syntactically very different: the first takes a *sentence* (or open sentence) for its complement; the latter takes a subsentential expression. Permit to explain my meaning. (A) is famously ambiguous between de dicto and de re readings (Quine 1966). The de dicto reading is: John wants that there is some boat x such that John has x. According to Quine (1966), the de re reading is: There is some boat x such that: John has x. But if Quine is right, then “wants” sometimes denotes a relation to a proposition (de dicto reading) and sometimes denotes a relation to an object, and not a proposition (de re reading). But it is quite clear that, even in the de re case, what is wanted is the truth of a certain proposition. John does not simply want a *boat*. There is some specific boat x such that what John wants is: that John has x. So the de re reading is: There is some boat x such that John wants that John has x. What is wanted is always the truth of a proposition, even in de re cases of wanting. In the de re case, there is some specific object x such that one wants the truth of a proposition of the form:…x…In the de dicto case, there is no such object; the truth of the proposition in question doesn’t turn on facts about this or that specific boat.

But one touches an *object*, not a proposition. So “wants” denotes a relation between a person and a proposition; “touches” denotes a relation between an individual and an object. So a case can be made that, when we make perspicuous the syntax of (A) and (B), it turns out that (A) has the form *for some x, aRx*, while (B) has the form *a wants that: for some x such that phi(x), a has x* or *for some x such that phi(x), a wants that: a has x.* So (A) and (B) *do* have different syntactic structures. Many linguists hold, for more or less the reasons just given, that (A)’s *syntactic* form – not simply the form that some logical reconstruction of it has – is different from (B)’s Similar remarks favor the view that (1) “John knows that snow is white” and (2) “John thinks that snow is white” have different syntactical structures, and that this difference in syntax is responsible for the different logical properties of those sentences

(e.g. that the first implies that snow is white, while the second does not). For related remarks, see Chomsky 1975 (93-95).

The point I am making is that syntax is so deeply embedded in semantics that, until we have fairly high resolution knowledge of semantics, we cannot make judgments about the extent to which syntax and semantics line up; and we cannot make syntactic judgments at all. See Chomsky 1975 (95). So, even when we are dealing with "perspicuous" formal languages, sensitivity to syntax presupposes fairly high-resolution semantic-knowledge.

In many places, Chomsky has argued compellingly that (1) "John expected Bill to leave" and (2) "John persuaded Bill to leave" have very different syntactic structures, even though, from the viewpoint of the grammar we learned in school, they have the same syntactic structure. (See Chomsky 1965.) This bears out our position.

Our position is further confirmed by consideration of the problematic relationship between phonetics and semantics. Sentences of natural language are typically replete with phonetically unrealized constituents. Here is an extreme example. You ask Bob "who killed Caesar?". Bob says: "Brutus". The naïve and false thing to say is:

"The semantics of Bob's utterance was a proposition of the form *aRb*. But the syntactic form of Bob's utterance was simply *a*. So Bob's utterance involved a misalignment of semantics with syntax."

The first sentence is correct. But it is easy to make a case against the second two sentences. Bob's utterance is plausibly seen as elliptical for – as a mere phonetic condensation of -- "Brutus killed Caesar." If that is right, then that utterance is simply an abbreviated – a phonetically condensed – way of saying "Brutus killed Caesar". From the viewpoint of semantics, though not of phonetics, Bob *did* say "Brutus killed Caesar". So when we undo the distortions of phonetics, we find that Bob's utterance *did* have the form *aRb*. In general, when we undo phonetic distortions, we find, arguably, that there is an alignment of semantics and syntax. I suspect that our folk-syntactic theories embody a failure to undo the distortions of phonetics; they put too much syntactic stock in what are really just methods of abbreviation that have no syntactic significance (e.g. saying "Brutus" instead of "Brutus killed Caesar"). In my view, that is why folk-syntax (the grammar we learn in school) is so out of alignment with semantics. Folk-syntax lumps together (A) and (B) because in natural language, for reasons of efficiency, important syntactic differences are telescoped, hidden beneath phonetic condensations, and we naively mistake the chaff of ellipsis for the wheat of semantics.

At one point in the *Tractatus*, Wittgenstein says that, where natural language is concerned, "everything is in order". I myself do not know quite what he means. But one possible interpretation lies in the points just made. The Russell-Frege view is that sentences of natural language are inherently obscure: there is a divergence between what they *seem* to say and what they actually say, and this divergence owes to there not being a perfect alignment, in natural language, of the syntactical with the semantic. The point I was making in response is that what is at fault is not natural language, but our crude theories about how natural language functions – theories that are embodied in the grammar we are taught at school. Because these theories are so approximate (because, for example, that theory puts "some man" and "John" in the same syntactic category), we end up believing that the "syntax" of natural language is out of alignment with its semantics. But what is out of alignment is our *theory* of the syntax of natural language, not the syntax itself.

[168]

To block this line of argumentation, CTM must say this:

"Where symbols of natural language are concerned, any shape can be associated with any syntax at all, given the right semantic background. But things are different where Mentalese symbols are

concerned. In Mentalese, morphology *does* realize syntax."

But, as we saw in footnote 35, this line of thought is probably not available to CTM.

[169] Advocates of CTM often talk about how semantics can be "coded" in morphology; and they say that, because Mentalese semantics is appropriately "coded" in Mentalese morphology, morphology-driven operations on such symbols constitute content-driven operations. The word "coded" here is ambiguous; and neither disambiguation is favorable to CTM.

On one interpretation, to say that Mentalese semantics is "coded" in Mentalese morphology is to say that there is a correspondence – a one-one function (or, at any rate, some kind of function) – from Mentalese semantics to Mentalese morphology. As we just saw, in *this* sense of the term "coded", given only that Mentalese semantics is coded in Mentalese morphology, it does *not* follow that a morphology-driven operation on Mentalese symbols is a computation or inference of any kind. What *drives* the play of symbols must be a *knowledge* of that correspondence, a *knowledge* of that coding-relation; and such knowledge is obviously semantic (and, indeed, extra-semantic).

On another interpretation of the word "coded", to say that Mentalese semantics is "coded" in Mentalese morphology is to say that Mentalese morphology *is* Mentalese semantics or, in any case, that Mentalese semantics supervenes on Mentalese morphology. But, as we saw in footnotes 35 and 41, it is probably not open to CTM to say that Mentalese semantics supervenes on Mentalese morphology.

[170] Of course, where non-analytic statements are concerned, knowledge of the *extra*-semantic is also required.

[171] Or other physical entity, e.g. sound or pattern of light.

[172] I don't know whether, at the level of detail, this characterization is accurate. But my point is that applying that algorithm consists in making a judgment, in recognizing the truth of a proposition, indeed an extremely conditional one. It may turn out – probably will turn out – that the proposition in question is even *more* complicated and conditional than we have here said. But that would only confirm my point that to apply an algorithm, to compute, is to recognize the truth of a proposition, and is not to manipulate shapes.

[173] Even the use of truth-tables to perform a computation is not, strictly speaking, a mechanical process. For every language in which a given shape – say, "P or not P" -- expresses the disjunction of a proposition and its negation, there are infinitely many in which it expresses the *conjunction* of a proposition and its negation, and infinitely many in which it expresses nothing at all. The same is true *mutatis mutandis* of "if P, then P", "P and not P", "if P, then P or Q", and so on. *If* those shapes are taken to have certain meanings, *then* there is a function -- embodied in the method of truth-tables -- assigning truth or falsity to certain expressions depending on their shapes. So using truth-tables to perform a computation (to determine truth-values) thus presupposes semantic background knowledge; it presupposes a knowledge that certain semantic conventions are operative.

[174] See Coulter 1983 (57-58), Wittgenstein 1958, and also Wittgenstein 1983 (422), especially: "'Mechanical'—that means: without thinking. But *entirely* without thinking? Without *reflecting.*"

[175] If you are a talented pianist, you can automatically play (or even sight-read) a piece. If you are an extremely talented composer, you can automatically compose a sonatina. But these operations though

"automatic" in some sense, are not "mechanical"; such operations though executed automatically obviously implicate a rich backlog of knowledge and of cognitive abilities.

[176] Again, it may be appropriate to reference Wittgenstein 1983, especially pages 257-258 and 422.

[177] Compare Wittgenstein 1983 (257).

[178] At least not in virtue of the fact that it is manipulating shapes: there is always the possibility that, for some entirely unrelated reason, it is engaging in cognitive processes.

[179] See Wittgenstein 1983 (257-258, 422). See Coulter 1983 (65-68).

[180] Suppose it turns out that thought *does* consist in morphology-driven operations on symbols. This is, I suppose, an epistemic possibility. Brain-research might reveal that a person is thinking *Socrates is bald, therefore something is bald* exactly if there occurs in that person a sequence of brain-states with certain morphologies. (This is not my own belief; but I do not wish to deny, at least not in this context, that this *might* turn out to be the case.)

But if it turned out that thought consisted in purely morphology-driven sequences of brain-states (a very dubious empirical proposition), that would not (at least not in any direct way) vindicate the idea that thought is fundamentally *computational.* For a computation is not a manipulation of symbols; a computation is *not* a morphology-driven sequence of symbols. A computation is a *judgment* regarding the validity of strings of symbols. Such a judgement may, or may not, result in a manipulation of shapes.

[181] "Psychophysical and theoretical identifications" *Australasian Journal of Philosophy* 50 pp.249-258.

[182] See Fodor's introduction to his collection of essays *Representations* (MIT 1983) for a clear statement of what functionalism is and also what the merits of that doctrine are. What I say in the next few paragraphs are very clearly stated there.

[183] I myself don't have this intuition; neither does Chris Hill. (See *Sensations*: Cambridge 1991.) I think it perfectly possible, at least epistemically, that to have beliefs, tickles, and so forth, one has to composed of quite specific kinds of tissues. I would actually be quite surprised if, to have sensations, it were *not* necessary to be composed of quite specific sorts of tissues – I would be very surprised if creatures made of aluminum could feel anything.

[184] See the last footnote. I am not stating my own reasoning. I am simply saying why *in general* type-type materialism is rejected.

[185] Of course, this example is grossly simplified. For *any* mental type T, it is extremely difficult to state the causes and effects that a thing x must have, from a functionalist's perspective, to qualify as instance of T. This in itself, I think, is a problem for functionalism: For *no* type T have I ever seen anyone say, except very schematically, what causes or effects x must have to fall under T. And this raises the question: whether there *are* any such causes and effects. But this cannot be pursued here.

[186] Sensations, of course, can have purely physiological causes, e.g. a kick in the shin, a grain of sand on one's eye. But among their characteristic *effects* are mental states, e.g. a desire to remove the grain of sand, a desire to take an analgesic. So even those mental entities that have the thinnest characteristic causal connections to other mental entities *still* have such connections

[187] It isn't hard to hypothesize circumstances under which you might need to rely on a theoretical apparatus to learn that snow is white this would be the case if your belonged to a species of very large

or very small creatures who couldn't *see* snow. Even if you just happened to live some place where there wasn't snow, so you'd have to *figure* out that snow was white.

[188] Stich's point is not to prove that functionalism is false. It is to prove something else entirely – that 'folk-psychology' is false. Stich wants to show that the categories in terms of which most psychological explanation is couched (*belief, desire,* and so forth) are explanatorily bankrupt.

[189] Stich's emphasis. *From Folk-Psychology to Cognitive Science: The Case Against Belief* Bradford Books, the MIT Press, 1983, pp. 25-26.

[190] This doesn't mean that functionalism is wrong. But it means that, if we accept Lewis's way of making functionalism be non-circular, then we have given up on the idea that functionalism vindicates any kind of materialism. For, as we just noted, Lewis makes functionalism be non-circular only at the cost of having *uneliminated* mental terms.

[191] "The Foundations of Mathematics" in *Philosophical Papers* (D.H. Mellor ed.) Cambridge: 1990.

[192] I should point out that Russell (in "Knowledge by Acquaintance and Knowledge by Description") held that some propositions are ungraspable: any proposition that has an external object (as opposed to some description thereof) for one of its constituents is, for Russell, ungraspable. Contemporary direct reference theorists sometimes hold that there propositions we cannot grasp, e.g. propositions that are object-involving with respect to future individuals.

[193] Of course, the inference may be inductive or deductive.

[194] Although the word 'concept' is ambiguous, that ambiguity is not as radical as that associated with words like 'bank' and 'rent'. Presumably, a mental entity qualifies as a concept_O in the psychological sense – as a *grasp* of something – if it somehow encodes a concept_O in the *objective* sense. I grasp the concept_O of triangularity – which is a non-mental entity – by virtue of having a mental content which somehow encodes that concept_O. So a concept in the psychological sense is a mental entity which encodes a concept_O in the objective sense.

Also both concepts and conceptso are *modes* of presentation. The concept_O *class of all points equidistant from a given point* is a mode of presentation of the property of sphericity, and so is the concept_O *closed three-dimensional figure of uniform curvature.* So underneath the superficial ambiguity of the word "concept", there are identities of meaning.

[195] Of course, you could think about your grasp of Socrates. But what enables you to think about Socrates is that you *have* a concept of Socrates, not that you *grasp* your concept of Socrates. You don't have to grasp your own concept_O – your own grasp -- of Socrates to think about Socrates.

[196] I think that intentionalists might say that the very ability to have the phenomenology involves a background of representational abilities.

[197] Fodor (1998:13).

[198] Fodor (1998:35).

[199] Fodor (1998: 80-86).

[200] Fodor 1981e.

[201] Fodor (1981e:283-292).

[202] Fodor (1981e:283-292).

[203] Fodor (1998:46-47).

[204] See Kuczynski 2004b, where the following argument is given in much more detail.

[205] The reason is that we give "wide scope" to the descriptive information. I tell you "some guy stole my wallet." You then think to yourself: *he* didn't have to become a thief – *he* might have become a police officer. Here you are *not* thinking: *it is possible that somebody x uniquely stole __'s wallet and also that x did not become a thief.* What you are thinking is: *somebody x uniquely stole __'s wallet and there is a possible circumstance where that x did not become a thief.* So the descriptive information is given *wide* scope with respect to the modal operator. Even though it is always through descriptive information that we acquire concepts of objects (or so I will argue), that doesn't corrupt our modal intuitions; for we give the descriptive information "wide scope" with modal operators.

[206] Or *seemed* to argue: see below

[207] Jerry Fodor (1998) has gone so far as to say that a causal connection to x is *all* that is involved in having a concept of x. So I can have *no* beliefs about the number and yet have a conception of it: I need only have a casual connection to it (or, more accurately, some instance of it). I need only have a causal connection to Socrates to be able to think about him. I needn't know *anything* about him. My book__ is a discussion of Fodor's atomism and his arguments for it.

[208] It is possible that what Kripke says is not so much that people know of *no* individuating descriptions of the things they think about and refer to. His point is that the individuating descriptions are not the ones that Russell (and possibly Frege) *thought* they were. I think that Kripke himself says something similar to what we say below – when we argue that one *does* need an individuating descriptions *of a very special kind* to refer to an object. (I won't press these exegetical points. I will argue only for the point that none of Kripke's cogent points to any degree warrant the rejection of the "descriptivism" I have been urging.)

[209] What I will say goes a bit beyond what Kripke says; my own commentary is interlaced with an exposition of Kripke's views.

[210] More exactly, that person gives Socrates the Ancient Greek equivalent of "Socrates", whatever that is. In this discussion, we will ignore the fact that, over the years, names may undergo considerable phonemic change. Obviously "Confucius" sounds very little like the name of the old sage. These phonemic changes introduce no philosophical mysteries, but allowing for them would clutter up the exposition considerably.

[211] In fact, it is arguable that *all* such descriptions do, except when the description in question refers to some abstract object, like the number two. The (often hidden) presence of indexicals is needed to prevent problems relating to what Strawson called the "massive reduplication". Your friend Fred says to you "somebody uniquely invented bifocals; his name is 'Ben Franklin'". As it happens, on a distant planet, some other person (or intelligent creature) also invented bifocals. Is Fred *wrong*? Not if we take him to mean: in *this* context – in the context of the world *we* occupy – somebody uniquely invented bifocals. Surely we would not take Fred to rule out the possibility that on some distant planet there were

intelligent, bifocal-inventing life-forms. So, in effect, we attribute a hidden indexical to Fred's statement. For reasons of this kind, we must see our discourse as permeated by indexicality, hidden though it may often be.

[212] Of course, the term "occasion" has to be defined somewhat narrowly. One can refer to twenty different people in a single conversation and thus on a single "occasion". But at any *specific* juncture of that occasion – at any *one* time (leaving aside ambiguity) – one is referring to just one person or object. When I say "occasion" I mean a specific juncture of a discourse.

[213] See Russell 1948 (89-99).

[214] Russell may well have chosen historically pregnant examples for purely illustrative reasons.

[215] It isn't even clear if Russell *was* wrong about the nature of the uniquely individuating descriptions. It could be that he was simply *using* historically pregnant descriptions to illustrate his basic point, which is that one must have *some* uniquely individuating description of Socrates to think about, or refer to, Socrates. If he had illustrated his point with some uniquely individuating description that didn't *obviously* apply to Socrates – e.g. *the guy to whom Seymour was referring on occasion C.*…--nobody would have followed what he was saying. So his use of uniquely individuating descriptions like *the philosopher who drank Hemlock* may not really indicate his belief that *that* is an actual uniquely individuating description through which one might refer to, or think about Socrates. It might just be a superficial expository device. But I am not a Russell scholar; so I remain open on this.

[216] I am prescinding from the controversial question whether there is a difference between "conceptual" and "non-conceptual", or between "propositional" and "non-propositional" content. (I discuss this issue at length in__) It is clear that, whatever the right resolution of this dispute may be, the content of your perceptions is, at least, in part, existential: you see *a situation with such and such properties.*…You see *a* creature next to *a* chair…

[217] In fact, you don't really *see* that it is Fido. Given only the information embodied in your sense-perception, what you are seeing could be twin-Fido or Android-Fido. Your judgement that you are seeing Fido is meta-perceptual; it involves background knowledge and plausibility-considerations, and so forth.

[218] Kant, Brandom, and others have stressed that the veritable objects of mental representations are truth-evaluable contents: things that correspond more closely to sentences or whole propositions than to nouns, adjectives, and the like. Even if you are just hallucinating, or seeing, *an apple*, the content of your state is not just: *an apple* – it is more like *there is an apple of such and such color in such and such place.*…The objects of *linguistic* representation are often mere *objects:* "Socrates, "Plato", and so on. But it would be unwise to suppose that there existed such a perfect parallel between facts about public systems of communication and facts about conception. Even though, semantically, "Plato" is a simple, it is far from obvious that my *concept* of Plato is correspondingly "atomic" (though Fodor 1998 argues just this. In __ I counter-argue.)

[219] I am not convinced that this is true of one's grasp of objects *in general.* For, at least arguably, it would be viciously regressive to say that one grasped *everything* this way. After all, to grasp a description, one must grasp its constituents; and if one had to grasp each of these constituents by a description, we'd obviously be launched on an endless regress. So it seems to me (as it did to Russell, for nearly enough the same reasons) that we grasp something "directly" (non-descriptively). It also seems to me (as it did to Russell, for nearly enough the same reasons) that certain states of one's own consciousness and at least some universals fall into this privileged category. But obviously the question

"*which* things do we grasp non-descriptively?" is a large one, and deserves a discussion of its own.

(It should be observed that Sellarsian anti-foundationalism might suggest that, in fact, *everything* is grasped "descriptively" – that there are no unmediated, non-descriptive graspings. So perhaps the line of thought just put forth is naively foundationalist. If that is the case, then it *would* be generally true that, to think about object O, one must know some uniquely individuating description that applies to it. Of course, if one counts trivial descriptions like *the unique x such that x=O*, then perhaps it follows trivially that one must know such a description to grasp O. But here I am leaving aside descriptions of that sort – descriptions that, to be understood, *presuppose* a grasp of the thing in question.)

[220] I say "may" for a couple of reasons. After one learns that (so to speak) Hesperus=Phosphorous, then *Hesperus has phi* may end up coinciding in cognitive value with *Phosphorous has phi*. (Of course, this point has to be adjusted for speakers of other languages.)

Also, depending on how one initially learns the meaning of "Hesperus" and "Phosphorous", *Hesperus has phi* *Phosphorous has phi* might never have had different cognitive values (for that person) to begin with. This will be the case if am introduced to the words by somebody who points to Venus and says: "*that* is called both 'Hesperus' and 'Phosphorous'".

[221] Let me end by considering an objection to this last line of thought (I will put the objection in the mouth of an imaginary locator):

"You have *ex nihilo* produced this strange theory of conception as being sustained by knowledge of interlocking existence claims. And you claim that recognition consists in seeing that some one object satisfies two existence claims. But all of this seems very *ad hoc* – something you've confabulated to get your bankrupt theory off the hook."

I think that this comment has force only to the extent that we've been jaded by wrong theories of sense-perception and, more generally, of objectual awareness.

When you see Bob, embedded in your sense-perception is existential information. You cannot get rid of that existential information without eliminating the perception. This suggests that that existential information is constitutive of the content of the perception. Granted, Bob himself is (part of) the representational content of that perception. But he is given to you via this existential information.

So when you see Bob on some occasion t, the content of that perception is given by some bit of existential information like: *there is some chap whom I'm now seeing, wearing a bowler hat…*

You see Bob again, on occasion t*. Again, the content of your perception is given by some existence claim: *there is some chap I'm seeing, wearing a baseball cap…*

On both occasions, the content of your perception is existential.

Suppose that, on t*, you *recognize* the person you are seeing is the person you saw on t. Your t-perception is given by existential information, and so is your t*-perception. So if, on t*, you *recognize* that you are seeing the same person you saw on t, that must consist in your realizing that the same entity satisfies both existence-claims are satisfied – that there is some one entity who satisfies both. Once a few platitudes are granted about sense-perception – that sense-perceptions have existential content -- our theory of conception and recognition proves to be far from *ad hoc*.

[222] See 'Emotivism' in *The Cambridge Dictionary of Philosophy*.

[223] See *'Moral Realism'* in *The Cambridge Dictionary of Philosophy*.

[224] See Ayer [1], Chapter vi, Blackburn [1], Chapter 6, and Stevenson [1].

[225] Some qualifications are in order. First, in this paper, by 'meaningful', I mean 'literally

significant'; I mean, essentially, 'is such that, by virtue of its literal meaning, it is true or false.' So I am using the term 'meaningful' in a narrow sense. I will, if I am not mistaken, be using the term 'meaning' and 'meaningful' to denote what used to be called *cognitive* meaning. No one denies that valuative sentences have *a* kind of significance: they are surely significant *indicators* of the mental states of those who utter them. What is in question is whether they are significant in the sense in which 'Socrates was bald' is significant; what is in question is whether they encode propositions – whether they are significant at the level of literal meaning. Any sentence that is significant at the level of literal meaning is a conventionalized way of expressing a potential belief. 'Snow is white', '2+2=4', '2+2=5' are conventionalized ways of expressing things that could be the object of belief. That is why such sentences are literally significant. It might be said: *But "I do not exist" does not express a proposition that could be the object of belief, but it is literally significant*. Fair enough: but the *negation* of what is expressed by "I do not exist" could be believed. And it is clear that 'I do not exist' has literal significant precisely because what it expresses can be *dis*believed. So it would be fair to say that a sentence S is significant if it is such that it is a conventionalized way of encoding a potential belief *or* disbelief. In this paper, our primary focus is not the nature of literal meaning. I think it would lead to unnecessary complications and obscurity if I repeatedly said 'a sentence is a conventionalized expression of a belief *or* a disbelief'; so I will just say – even though *technically* it may be inaccurate – 'a sentence is significant if it is a conventionalized way of expressing a belief.' Also, a *dis*belief is surely a kind of belief. So it *may* turn out to be *technically* accurate to say that a sentence is meaningful if it is a conventionalized way of expressing a belief.

Some may balk at my claim that anything that is a conventionalized way of expressing a belief/disbelief is a *sentence.* "If I use smoke signals to indicate that I am in danger, have I produced a *sentence*?" My answer is: *yes.* You have a produced a sentence that belongs to a language that, instead of using phonemes, uses smoke-signals. There is, ultimately, nothing to being a sentence *other* than being part of an artificial and systematic means of encoding propositions and, therefore, of expressing beliefs/disbeliefs.

In any case, whatever complications may connect with my analysis of literal meaning, I would be very surprised if these really undermined what I have to say about the relationship between emotivism and moral realism. To be sure, what I say about that relationship may be wrong or under-argued – but not in virtue of any deficits in my discussion of literal meaning.

[226] There is also the very real psychological possibility that, although they rejoiced in their evil deeds, Bundy and Dahmer – and other serial killers – *at some level* did have negative emotions towards them. The self-destructiveness evinced by many – though by no means all – serial killers is often chalked up to their own (unconscious) disgust at their deeds. There *does* seem to be the possibility, then, that if a serial killer sees an act that he committed as wrong, that is because, at some level, he *does* have the kind of emotion towards that act that, according to the emotivist, is expressed by sentences of the form 'x is wrong.' There is a striking anecdote about Ted Bundy. Before he was unmasked as a killer, he once asked a friend of his where in the United States a murderer was most likely to be put to death. The answer he was given was Florida. Bundy committed his final murders, was caught, and executed in Florida. Of course, these considerations do not show that it is not *possible* for someone to see something as wrong without having a negative emotion in connection with it. I bring up this point simply keep this highly abstract ethical discussion in contact with some psychological realities or, at any rate, possibilities.

[227] This exact statement was made to me by Ernan McMullin in response to a lecture I gave at Marquette.

[228] Sartre was a cognitivist: 'Emotion is a certain way of apprehending the world…' (in Calhoun and Solomon: 246) And so is Robert Solomon: 'My emotions *are* judgments I make…' (Solomon's

emphasis) (Calhoun and Solomon: 316). See also the section entitled 'Affectivity' in Sartre's *The Psychology of the Imagination.* The argument I set forth below for cognitivism is quite similar to Solomon's. And I think that, ultimately, notwithstanding certain superficial dissimilarities, it is not so different from Sartre's argument. Ultimately, all arguments for cognitivism devolve on considerations like those set forth below, though there is of course some variation at the level of detail.

If I am not mistaken, the entire 'phenomenological' school of thought endorses, and may be committed to, some kind of cognitivism about emotions. Cf. Merleau-Ponty *The Phenomenology of Perception.* And among analytic philosophers, cognitivism is one of the dominant views (though it may not be *the* dominant view). Cf. Calhoun and Solomon *What is an Emotion?* (Oxford 1993) especially the introduction.

[229] Of course, some people will say that there cannot be object-less moods. Indeed, some people say that *everything* mental has an object. But such people, I think, are forced to say that 'mere' moods and feelings are really emotions. This of course does not mean that emotions *are* beliefs. But it means that we should take seriously the idea that they are beliefs and, therefore, that we should take seriously the implication I am drawing from 'emotions are beliefs' to 'emotivism is a form of realism.'

[230] Calhoun and Solomon: 246. Sartre's emphasis.

The paradigm anti-cognitivist is Hume: "A passion is an original existence…and contains not any representative quality, which renders it a copy of any other existence or modification. When I am angry, I am actually possest with the passion, and in that emotion have no more reference to any other object, than when I am thirsty, or sick, or more than five foot high. 'Tis impossible , therefore, that this passion can be oppos'd by, or be contradictory to, truth and reason; since this contradiction consists in the disagreement of ideas, consider'd as copies, with those objects, which they represents.' (*Treatise* p.414) It is obvious, at the very least, Hume is quite wrong to say that (e.g.) anger contains *no* 'reference to any other object': I am angry, I am angry *at* Bob or Sally or Fred. There is inevitably a 'reference' to an object. In *What is an Emotion?* (Calhoun and Solomon, editors) numerous attacks are waged on the Humean position.

[231] This is Thomas Nagel's way of putting it. See 'What is it like to be a bat?'

[232] The English word 'feeling' is, I think, ambiguous between *phenomenally pregnant state* and *attitude.* I think that a similar ambiguity occurs in other languages: the verb 'sentir' has, I believe, a corresponding ambiguity. It is understandable *why* there is this ambiguity: presumably it stems from the fact that if one 'feels' a certain way about something in the sense that one has a certain *attitude* towards that thing, then one *tends* to have certain *feelings* – in the sense of *phenomenally pregnant mental state* – in connection with that thing. There is *some* kind of connection between e.g. hating someone – which is non-occurrent and, in and of itself, has no phenomenal content – and having certain negative 'feelings' in the narrow, purely phenomenal sense of this word. (The exact nature of this connection is unclear and has been the subject of heated debate.) But, for our purposes, it is enough to know that the word 'feel' is ambiguous between these two senses and that I am using this word to refer to *attitudes*, not phenomenally pregnant states (except in so far as some attitudes happen to be phenomenally pregnant).

[233] See Blackburn [2]. .

[234] See footnotes 2-4.

[235] Fodor, Jerry *The Language of Thought* Thomas Y. Crowell Company Inc. 1975 p. 27

[236] Tarski, Alfred 'The Concept of Truth in Formalized Languages' in *Logic, Semantics,*

Metamathematics Hacket 1983

[237] If externalism is correct, then the meaning of a word is not (always) *just* determined by what people think, but also by the nature of the external environment, the history of that word's employment, etc. At the same time, in all cases what a word means is generally determined *in part* by the cognitive processes surrounding employment of that word: a 'word' used by a zombie is no word at all: even the most hard-line externalist would probably have to concede that much.

[238] Fodor [1].

[239] See Horst [1].

[240] See Fodor [1].

[241] Some authors use the word 'concept' to refer only to the ability to have certain kinds of *abstract* thoughts – to, for example, the ability to think about redness *in isolation* of red objects. I will *not* be using the word 'concept' in this restricted sense: in my usage, one's ability to think about a particular person or a particular rock is a 'concept' – and so is one's ability to think about personhood and rockhood in the abstract.

[242] See Fodor [2].

[243] This is an old point. I think it was first defended by Wittgenstein. See Wittgenstein [1].

[244] See Sellars [1].

[245] Or, presumably, an animal's.

[246] Some clarifications are in order. First of all, when I talk about the mental contents which a person has, I don't mean the ones he has at a particular moment; but the totality of mental contents he has throughout his lifetime. Obviously a person cannot be identified with the mental contents which he has at some particular moment.

Second – and probably more importantly -- I do not want to say that, according to the bundle-theorist, a person is strictly speaking *identical* with the totality of mental contents which he has. For, presumably, the person has modal properties which the class of mental contents which he has does not. That class is modally frozen: if it were to contain a different member, it would ipso facto be a different class. By contrast, a person could obviously have different contents from the ones he has. (Compare: the statue is not *identical* with the clay of which it is composed. For the statue and the clay have different modal properties. The destruction of the statute doesn't necessarily involve the destruction of the clay.) So a person cannot strictly speaking be *identified* with the class of contents which he has throughout his life. But what *can* be argued for is that a person is wholly *constituted* by the totality of mental contents which he has throughout his life. It is *this* view that I shall refer to as the 'bundle theory'.

[247] Hume of course was a great proponent of the bundle theory. (Indeed, he may – though I'm not sure about this – have invented it.) 'All [of one's mental contents] are different, and distinguishable, and separable from each other, and may exist separately, and have no need of any thing to support their existence. After what manner, therefore, do they belong to the self; and how are they connected with it? For my part, when I enter most intimately into what I call *myself*, I always stumble on some particular perception .When my perceptions are remov'd for any time, as by sound sleep; so long am I insensible of *myself*, and may be truly said not to exist...I may venture to affirm...that [people] are nothing but a

bundle or collection of different perceptions.' Hume [1] p.252.

[248]

See, for example, Carruthers [1] pp. 54-57..

[249]

There is another, perhaps a more perspicuous, way to state what I am trying to do. If (i) is true, then selves are built out of mental contents, just as building are built out of bricks: a self can no more exist independently of the mental contents that belong to it than a building can exist independently of the bricks that make it up. But a given self is not just built out of just any mental contents: it is built out of mental contents that *cohere* in a certain way, that stand in a certain relation of interdependence. Our building metaphor might illuminate this point. We cannot say of *any* two bricks that they are constituents of a single building; we can say this only of pairs of bricks that cohere in a certain way, that are physically related to each other in a certain way. Analogously, we cannot say of just *any* two contents that are constituents of a single person; we can say this only of two mental contents that cohere in a certain way: that stand in a certain coherence-relation to one another.

Many have said that there is no non-circular way to define this coherence-relation. What is essentially equivalent: that, with regard to the coherence-relation that holds between two contents belonging to a given person, in order to define that relation, we must invoke the idea of the self to whom both those contents belong. If this is correct, then selves are more basic than the coherence-relation in question. In that case, of course, selves are more basic than bundles of mental contents, and therefore cannot be identified with bundles of mental contents. As against this, I will try to show that we *can* give a non-circular definition of the aforementioned coherence-relation and that we *can* therefore rescue the idea that selves are wholly constituted by the mental contents which (as we say) they 'have'.

[250]

[251]

By *purely* physical interactions I mean physical interactions that do not have a mental component. By no means do I wish to suggest mental interactions are not themselves physical interactions. (I wish to remain neutral, at least in this context, on the question whether materialism or dualism is correct.)

[252]

Before proceeding, I should make one point explicit. I have said that, when we are dealing with the interaction of mental entities, we can speak (in a literal, non-metaphorical way) of the outcome of such an interaction as being *appropriate* to the mental entities involved. (My visual perception of Mary holding a chainsaw and my knowledge that Mary wants to kill me in some sense 'combine', and thus interact, to form an intention to flee from Mary. Surely the resulting intention is, in a non-metaphorical, non-artificial sense, *appropriate* to the aforementioned perception and knowledge.) So mental entities seem to have a normative (or, as I will henceforth say, *thematic*) *content*. We can, non-metaphorically, say that a given mental entity M would appropriately occur in such and such a context. In saying that mental entities have *this* sort of content, I am not saying that they have *representational* content (roughly, that they have conditions of satisfaction – that they are non-derivatively true or false). It is a hotly disputed question whether all mental entities have representational content. But, if what we've said is correct, it is relatively uncontroversial to suppose that they all have a certain kind of content – what I am calling 'normative' or 'thematic' content. For (so it seems) any given mental entity is non-derivatively appropriate in certain contexts (and thus non-derivatively appropriate) in others: *ipso facto* mental entities non-derivatively have a normative dimension or, as we might also say, normative (thematic) *content.*

[253]

This definition of what a self is may seem unable to survive thought-experiments like the following. Suppose that, one day, Joe 'divided' into two people A and B (just as amoebas divide into two people), and A and B were equally well connected, both psychologically and physiologically, to

Joe. It is said that Joe couldn't' be identical with both: for if he were, then given that identity is a transitive relation, it would follow that A was identical with; and it is clear that A is not identical with B. However, A's mental contents are I-related to Joe's mental contents, and so are B's mental contents. So our analysis seems to identify A and B.

How are we to respond to this? First of all, this kind of scenario poses a problem for *every* criterion of personal identity; it poses just as much of a problem for those who espouse a bodily theory of personal identity as it does for those who espouse a Humean theory; and it is certainly problematic for those who hold that a and b are identical in virtue of sharing a common 'transphenomenal' or 'transcendental' ego. So this kind of consideration is no more problematic for our theory than it is for any other.

But for argument's sake, let us concede that this kind of consideration disproves our theory. Even granting this, we have still, I think, given a satisfactory analysis of what it is for two mental contents to belong to a single person-*branch.* What do I mean by a 'person-branch'? Suppose that Joe splits into A and B at time t. Consider the four dimensional object composed of Joe up until t and of A after than: this object is a single person *branch.* (So therefore is the object composed of Joe up until t and then of B after that.) Our analysis does, I think, give a satisfactory analysis of what it is for two mental contents to belong to a single person-branch; and therefore of what a single person-branch *is.* (A person-branch may be defined as the class of mental contents I-related to a particular mental content.) Of course, in actuality, people-branches always coincide with people. So our analysis of personal identity is at least *extensionally* correct.

There is an altogether different class of objections to any Humean (or, more generally, any mentalistic) analysis of personal identity. Bernard Williams and others have set forth thought experiments that suggest, if they do not actually prove, that personal identity is to be analysed in terms of purely *bodily continuity.* There simply isn't room enough here to consider this type of objection. (See Williams [1].) But *prima facie* it would seem much more plausible to suppose that personal identity is to be analysed in terms of mental continuity, and not in terms of bodily continuity (except ins o far as, of course, mental continuity is a special case of bodily continuity); so it is certainly worth our while to try to come up with a reasonable mentalistic criterion of personal identity. In particular, it seems especially to come up with a mentalistic criterion of personal identity that is *Humean* in character. For any non-Humean, mentalistic criterion would, presumably, assume the existence of some kind of transcendental ego; and there is some reason to believe that the very conception of such an entity is an incoherent one.

[254] See Fodor’s introduction to his collection of essays *Representations* (MIT 1983) for a clear statement of what functionalism is and also what the merits of that doctrine are.

[255] It is a matter of great controversy whether the representational content of a perception can be embodied in a sentence – whether e.g. the representational content of your perception can be fully captured by some locution like ‘stop-sign of such and such color thus such distance away…’ Purely for expository reasons, I am pretending that the representational content of your perception of the sign can, at least in part, be embodied in a proposition. (Incidentally, this is not an unreasonable assumption: one can always *say* what one has seen, at least up to a point: ‘I saw a man in red sweater…’) But what is *not* controversial is that, whatever causal properties the perception has, it has these at least partially in virtue of bearing a certain representational content. And *this* is the important point in the present context. Whether that representational content can be embodied in sentences/propositions is a secondary question. And I am only pretending that it can be so embodied in order to make the overall outlines of my argument more perspicuous.

[256] See, for example, pp. 17-18 of *Representations*, where Fodor flatly states that he is a functionalist about everything except phenomenal content, and that phenomenal content cannot, in his view, be 'functionalized.'

[257] I would like to thank Rev. Ernan McMullen for reading this paper and for encouraging me to convert what was a lengthy footnote into part of the main text (now the last two pages of the present paper).

[258] A classical statement of this sort of foundationalism is found in Clarence Lewis' *Mind and the World Order* (New York: Charles Scribner and Sons, 1929). See also Laurence Bonjour's *Structure of Empirical Knowledge.* (Cambridge: Harvard University Press, 1998) and also Simon Blackburn's *Spreading the Word.* (Oxford: Oxford University Press, 1984).

[259] See Wilfred Sellars' *Science, Perception and Reality.* (New York: Routledge and Kegan Paul, 1963).

[260] See George Berkeley's *An Essay Towards a New Theory of Vision.* (Oxford: Clarendon Press, 1934). In that work, Berkeley still concedes that objects have mind-independent existences.

[261] It has been maintained, with some plausibility, that despite Kripke's insights, "Socrates" may semantically encode general sortal information. But, even if this is true, that information is surely too impoverished to count as a *uniquely individuating description* of the man.

[262] The asterices are quasi-quotes.

[263] Of course, one *could* say this:

Because you haven't been given any kind of definition of "Argo", you can*not* really use that term with understanding, and you cannot really know what is meant by sentences containing it. When you produce the sounds "Argo sounds like a bad person", you are really making some kind of metalinguistic statement, some statement like: " 'Argo' refers to somebody who is bad". And when you hear sentences like "Argo was a cad", you cannot really understand them; what you understand is some related meta-linguistic claim like: "'Argo' refers to somebody who was a cad".

This position cannot be summarily dismissed; and I will not dismiss it. But, for our purposes, it is enough to note that the objector's point *seems* quite false. Cases like the one just described are pervasive. There are many cases where no ostensive or descriptive definition is involved in our appropriation of a proper name. We hear the term being used; and – *presto!* – can use it ourselves. If the objector's point is correct, we are not really *using* any of these terms: we are merely *mentioning* them; when we produce sentences containing them, we are talking not about people, but about symbols denoting people. But this seems false. I have never been given a descriptive or ostensive definition of the term "Charlemagne". But it certainly *appears* as though I can make statements about Charlemagne, and not just about "Charlemagne". For reasons of philosophical modesty, I will operate on the assumption that this appearance is veridical.

[264] The asterices are meant to be quasi-quotes.

[265] Here I am basically repeating a point made, in different forms, by Kripke, Kaplan, and Evans.

[266] I further discuss this in the last footnote of the present essay. I have placed that footnote at the end, because some of the points that we have yet to make figure prominently in its content.

[267] This is a very rough way of putting it. Strictly speaking, nothing has *both* reference *and* descriptive content: indexical *tokens* have reference; indexical-types have descriptive content.

[268] See Soames (2005) for a discussion of these views.

[269] We have said that to have a conception of x is to know some true proposition of the form…*C*…, where C is a uniquely individuating description of x. There are, of course, cases where a causal connection to x is essentially involved. But, as we have seen, the causal connection is always embedded in knowledge of a proposition of the kind just described.

But there is an obvious problem with this view: a problem that doesn't threaten the things we've said, but must still be dealt with. I myself do not know who invented the fugue. Let O be that person, whoever it is. (To simplify discussion, let us suppose that exactly one person invented the fugue.) I certainly know *some* true proposition of the form:…*C*…, where C is a description that applies uniquely to O. I know that the following are true:

(i) Somebody x uniquely invented the fugue and he had at least some musical ability.
(ii) Somebody x uniquely invented the fugue and he lived before the 19th century.
(iii) Somebody x uniquely invented the fugue and x invented the fugue.

Obviously knowing *some* uniquely individuating description that applies to O isn't enough to know who O is. And it therefore isn't enough to know which proposition is semantically encoded in sentences like:

(iv) "O was a talented composer".
(v) "O died of scurvy."
(vi) "O incurred the wrath of his less talented contemporaries."

Remember what we said earlier about "Julius" and "Newman I". Unless one knows *which* individuals these terms refer to, one doesn't know which propositions are encoded in "Julius was tall" and "Newman 1 will be a genius"; one has a simulated, but not an actual, understanding of what is meant by such sentences. One has knowledge of a *description* of the proposition that is encoded in such a sentence, but one doesn't know which proposition is thus described. One knows that "Julius was tall" is true exactly if somebody x uniquely invented the zipper and x was tall. So it is clear that knowing who so and so is, and thus knowing the literal meanings of terms that refer to so and so, does *not* consist (merely) in knowing some uniquely individuating description that applies to so and so.

What else is needed? The short, very approximate answer is: To identify is to *re*-identify. To know who so and so is, one must know of *two* existence-claims that are uniquely verified by so and so and, further, one must know *that* some one person uniquely verifies them. Suppose you meet Bob on Monday. Since you have just met him, there is a sense in which you don't know who he is. If somebody asks you, "do you know who this is?", you must say: "I am afraid I do not." But, of course, in meeting him, you are uploading information that enables you to think about him – you are acquiring a concept of him. As we saw, this information is existential. Suppose that, a day later, you see Bob again. This time, there is a sense in which you know who you are seeing. As we saw earlier, your *recognizing* Bob consists in your knowing *two* existence claims of which he is the sole verifier, and

also in your knowing *that* he uniquely verifies those propositions.

That is the *short* answer to the question "what is it to know who somebody (or something) is?". Here is the long answer. Identification is a contextual notion; it is always *relative* to some background question, or relative to some body of knowledge, that one knows, or fails to know, who somebody is. Suppose you have been living next door to Bob for years. You know very little about Bob. He doesn't talk much. He politely waves at you when he mows his lawn. But beyond that, he is a mystery. Later you and your wife are at the beach. You see somebody in the distance. Mary, who has poor vision, asks: "do you know who that is?" You, have better vision, say: "yes: it is Bob, our neighbor." From some viewpoint, you know who it is that you are seeing: you can identify that person.

The next day, C.I.A. agents come to your house. They brusquely ask you: "Do you have any idea who your neighbor is?" You say: "I don't understand. He's just some guy who waves at me when he mows the lawn." They say: "Your neighbor is really the ringleader of an extensive terrorist organization. The government has been looking for him for years…"

When you saw Bob at the beach, you *did* know who he was, at least from one viewpoint. But, from some other viewpoint, you have *never* known who Bob is, at least not until you were briefed by the C.I.A. agents. So you did, and you did not, know who Bob was.

What happened here? As we saw earlier, having a conception of an external world object involves two things: having knowledge of existence claims of which that object is the sole verifier; having a certain kind of knowledge *about* those existence claims – more specifically, knowing that some one entity satisfies them all.

Whenever somebody verbally, or ostensively, indicates an object, some existence claim, of which that object is the unique verifier, is implicated. Mary points out Bob to you. You see Bob: in seeing Bob, what is given to you is some existence claim like: *there is somebody x to whom I am attending, and x is standing in such and such place, and has such and such appearance.*…When Mary asks you "do you know who that is?", what is really being asked is something like: *there is somebody x over there, to whom we are attending, and x has such and such features: do you know who x is?* And in being asked *do you know who x is?* you are being asked to identify a more comprehensive set of interlocking existence claims of which x is the unique verifier. Because your vision is good, you are able to do this: you are able to say something like *somebody y is uniquely our neighbor and x is identical with y.* (Of course, these are not the words you would use. But this would be the import of your statement.) In saying "that is Bob, our neighbor", you are conveying (though this isn't what your words literally mean):

(*) *somebody x uniquely lives on the left side of our house, mows his lawn frequently, goes by "Bob", and x THAT PERSON over there is identical with x.*

The only qualification is the person who you are now seeing on the beach – the person whom you can see sharply, and whom Mary can see only blurrily – is given to you through an existence claim: *there is somebody y to whom we are now attending.*…So (*) really amounts to this:

(**)*somebody x uniquely lives on the left side of our house, mows his lawn frequently, goes by "Bob", there is somebody y (over there, building a sand-castle) to whom we are now attending, and y is identical with x.*

Your background knowledge of Bob consists in your having knowledge of a set of interlocking existence claims. So your being able to identify the man in the beach consists in your being able to generate knowledge another such existence claim, and in your knowing of that new existence claim that it is satisfied by the same individual who satisfies the others (i.e. in knowing some one individual uniquely satisfies that new existence claim *and* the other ones). So *relative* to that background knowledge – to your knowledge of those existence claims – you know the answer to the question "who

is that?"

But relative to other bodies of background knowledge, this is not the case. You know that there were various bombings, in various cities. You know that somebody or other headed the organization responsible. So you know something like:

(^) Somebody x is responsible for such and such atrocities.

You also know that you have a harmless-seeming neighbor. So you know something like:

(^^) Somebody y lives in the house to the left, and y waves at me while he mows the lawn.

When the C.I.A. agents say to you "you have no idea who Bob is", the import of their words is: somebody x is your neighbor, and somebody y committed such and such atrocities, and you don't know that x=y. *Relative* to (^) you don't know the identity of the man who waves at you as he mows his lawn. But *relative* to some other existence claim, you *do* know the identity of the person whom you and Mary are seeing at the beach. To identify someone (or something) is to connect one existence claim to some other existence claim. Identification is a relative notion. Suppose that Bob is given to you through existence claim E1. You may be able to relate E1 to E2, i.e. you may know that someone uniquely satisfies both E1 and E2. But you may not be able to relate E1 to E3, i.e. you may not know that someone uniquely satisfies both E1 and E3.

There is a sense in which any three year old knows what water is. But there is also a sense in which any three year old (except for a few prodigies) do *not* know what water is. Suppose you point to water and ask little Timmy, "what is that?" He says: "that is water." He has correctly identified the substance in the glass. So, in that sense, he knows the identity of the liquid in the glass.

(w1) There is some substance x such that Timmy drinks x, bathes in x, and so forth.

Timmy knows that w1 is true. Timmy also knows that

(w2) there is some substance y such that y is in the glass being indicated.

Further, Timmy knows that

(w3) There is some substance x such that Timmy drinks x, bathes in x, and so forth; there is some substance y such that y is in the glass being indicated; and y=x.

So *relative* to w1, Timmy can identify the substance in the glass. But relative to other existence claims, this is not the case.

(w4) There is a substance z such that z has thus and such chemical structure.

Obviously non-prodigy Timmy doesn't know that:

(w5) There is a substance z such that z has thus and such chemical structure; and there is some substance y such that y is in the glass being indicated; and y=x.

So relative to (w4), Timmy doesn't know *what is in the glass.* There is no paradox here. Identification is relative to background knowledge. Being able to identify an object is being able to connect it to background knowledge. Timmy can connect the liquid in the glass to certain bodies of knowledge, but not to others. So, from some certain viewpoints, but not all, Timmy knows *what is in the glass.*

Being able to understand

(J) "Julius was tall"

involves knowing *who* "Julius" refers to. (Here we are using "Julius" in Evans' sense, as a descriptive name that refers to that person, whoever it might be, that invented the zipper.) Knowing who "Julius" refers to involves knowing not *one* existence claim that he satisfies: it involves knowing (at least) two such claims, and also knowing that some one person satisfies them both. Further, knowing who "Julius" refers to is contextual: relative to some existence claims, you might know it; relative to others, you might not. So the concept of knowing what a term refers to, and therewith the concept of being able to assign meaning to a term, is contextual: so, therefore, is the concept of understanding a sentence-token.

[270] The tie-shop example, and many of the lessons I extract from it in the following pages, are of course taken from Sellars (1963a: especially 144-149). Some of what I say goes beyond what Sellars (1963a) says or, perhaps, even implies. But I would like to think that what I say here does develop some of the inner meanings of that great work.

[271] The property of being perceivable is a matter of degree, and also of context. Therefore, the same is true of the property of being theoretical. What you can *see* is a function, at least in part, of what your background theories are. (This is a very Sellarsian point.) But, rough and ready though it is, the concept of a perceivable property is not without its utility; and the fact that it is subject to qualifications does not have any bearing in this context.

[272] Burge 1993.

[273] I realize I am splitting an infinitive. But the context licenses a departure from the rule (questioned my many a linguist) that infinites should not be split.

[274] Sellars (1963b: 336).

[275] Byrne (2001) makes a comment very similar to this to argue against some alleged counter-examples to intentionalism due to Peacocke. Byrne's use of this kind of attack works against *Peacocke's* counterexamples, but not against mine. There is little resemblance between my counterexamples and Peacocke's.

[276] Byrne (2001).

[277] Of course intentionalists might say: ' no one is mad enough to suppose that *different species* have the same kind of phenomenal content in response to a different kind of sound (experienced under a given set of circumstances). And, of course, no one denies that creatures that have different

phenomenal contents (in response to a given kind of sound under a given set of circumstances) can hear just as veridically, if not more so, than we. Intentionalism – at least any sane version of it – applies *within* species, not *between* species.'

But this concession eviscerates intentionalism. Intentionalism wants a metaphysically necessary relationship between the representational and the phenomenal; not a merely causal, contingent one. Obviously if intentionalism is not true *across* species, then a given bit of representational content *can* be realized by very different phenomenal contents, and there is no long any kind of reduction of the phenomenal to the mental, and thus no *explanation* of the phenomenal in terms of the representational (at least none along intentionalist lines). But the interest of intentionalism is surely not that it register the idiosyncrasies of species, but that it captures some kind necessary fact about the relationship between phenomenal and representational content. And if this relationship doesn't hold across species, then that relationship isn't necessary. And it probably – almost assuredly - *doesn't* hold across species.

[278] Wittgenstein was the great critic of the idea that sentences are meaningful in virtue of being associated with certain *objects.* See Wittgenstein 1953 (sections 94-134), and also Wittgenstein 1958.

[279] Of course, this is ironic, as Quine himself was one of the great foes of the linguistic theory. Later on we will discuss this irony at some length.

[280] Ayer 1952 (77): "The principles of logic and mathematics are true universally because we are never allow them to be anything else. And the reason for this is that we cannot abandon them without contradicting ourselves, without sinning against the rules which govern the use of language, and so making our utterances self-stultifying."

[281] The best discussion of the linguistic theory that I know of is in Pap 1958 (chapter seven). Another excellent discussion is in Bonjour 1998 (chapter 3).

[282] Quine 1966.

[283] Given in Bonjour (1998).

[284] I came up with this argument on my own, but I am fairly certain that someone else must have thought of it before me. For it is fairly obvious.

[285] Though it is original, it is reminiscent of Quine's argument.

[286] I am referring, of course, to *Two Dogmas of Empiricism* (Quine 1951). Dummett (1978) said that it is "the most important article of the last fifty years." And Boghossian 1997 (331) wrote the following:

"This is what many philosophers believe today about the analytic/synthetic distinction: In his classic early writings on analyticity…Quine showed that there can be no sentences that are true purely by virtue of their meaning, and those that aren't. In so doing, Quine devastated philosophical programs that depend upon a notion of analyticity – specifically the linguistic theory of necessary truth, and the analytic theory of a priori knowledge…I don't know precisely how many philosophers believe all of the above, but I think it would be fair to say that it was the prevailing view."

[287] For an excellent discussion of this issue, see Pap 1958 (chapter 6).

[288] Assuming that the substitutions are uniform and are significant: to the term "gladly" couldn't be substituted for "green".

[289] See ___ for an extensive discussion of this issue. (Reference deleted to ensure blind reviewing.) See Blackburn 1984 (214) for an elegant discussion of the epistemological problems associated with the possible worlds approach, along with some other problems (not mentioned here) that it has. See Peacocke (1999) for a clear statement of the epistemological problems associated with the possible worlds approach, and for an attempt to deal with them. Peacocke does not (in that work) address any of the other problems, here mentioned, for the possible worlds approach.

[290] See Jackson and Pettit 2004b-2004e, especially 2004c (48-49).

[291] See Kaplan 1968, 1989a, 1989b, and Putnam 1975.

[292] It would be more accurate to say that Smith's mental state was caused by *some state of affair involving Mary*. Strictly speaking, *Mary* doesn't cause anything, as we've noted.

[293] Perhaps one might distinguish *mental* content from *representational* content. One might say that Smith and twin-Smith have qualitatively identical mental contents, but different representational contents. There are two things to say. First of all, it seems at best arbitrary to separate so sharply between the mental and the representational. But suppose *arguendo* that we grant that distinction. In that case, by the argument just given, externalism strips representational content of causal power – not a palatable result.

[294] See Jackson and Pettit 2004b-2004e, especially 2004d (108-114) and 2004b (58-62).

[295] Things that are non-mental can in *some* sense be representational. The following deposit of ink – 'snow' – represents snow, even though that deposit of ink is non-mental. But it seems to represent snow only by virtue of its association with *mental* entities that represent snow. So the word 'snow' represents snow only *derivatively* of its connection with *mental* entities that represent snow. In general, whenever something non-mental is representational, it is that only derivatively of its association with something mental: with something that is *intrinsically* representational. In any case, this is the generally held view.

[296] Usually, the term 'intentionality' is used to denote what I am calling 'representationality'. But for reasons that will shortly become clear, this paper would become very confusing if I used the term 'intentionality'.

[297] A 'representational property' is a property such that, if a thing has it, that thing *ipso facto* has representational content; and something has 'representational content' if it is the kind of thing that *could* represent something – that has 'conditions of satisfaction.' This last point will be spelled out shortly.

[298] See Brentano [1].

[299] Dretske [1]. Quoted in Byrne [1] p. 204

[300] See Lycan [1]. Quoted in Byrne [1] p. 204.

[301] Quoted in Byrne [1] p. 200.

[302] It should be said that this doctrine has had its proponents: Brentano, Meinong, and others seem to have maintained that a hallucination does have a 'real' – in *some* sense of the word – object.

[303] Admittedly, some philosophers (e.g. Fred Dretske) have said that things like thermometers and weathervanes – things that we would tend to characterize as 'non-mental' – *are* non-derivatively representational; that the thermometer *really is*, in and of itself, saying that it is 35; that the

weathervane really is, in and of itself, indicating the direction of the wind. But if it is true that e.g. the thermometer *per se* is literally saying that it is 35 degrees outside, then it seems to follow that the thermometer *per se* has a kind of rudimentary mentality. And indeed, those philosophers who hold that thermometers, weathervanes, etc. are non-derivatively intentional, also tend to say that that have some kind of mentality. For example, Dretske says that, because the thermometer non-derivatively represents the temperature, there is *ipso facto* no categorial difference between it and (e.g.) and an indisputably representational, and therefore mental, entity like Joe's belief that it is 35.

[304] Actually, it would be more correct to say that propositions *are* representational contents.

[305] It must be pointed out that the word 'necessarily' here corresponds to *metaphysical*, as opposed to causal, necessity. To clarify this: The representationalist wishes to *reduce* phenomenal content to representational content. So if it is merely *causally* impossible, but still metaphysically possible, that two mental entities should have the same representational content while having different phenomenal contents, that would be inconsistent with representationalism: the mere *metaphysical* possibility that variation in phenomenal content might be independent of variation of representational content would be enough to show that there was something to phenomenal content over and above representational content. So, throughout this essay, whenever I mention any kind of necessity, it will be to metaphysical, not causal, necessity; when I use the terms like 'supervene' and 'necessary' and 'is wholly a function of', and the like, I will be using them in their metaphysical, not their causal, senses. (So when I say e.g. that x 'supervenes' on y, I mean that it is *metaphysically* impossible that there should be a change in x without a change in y. When I say that it is 'necessary' that a change in x correspond to a change in y, I mean that *metaphysically* – quite apart from the laws that happen to govern our universe -- a change in x must correspond to a change in y. And when I refer to the view that e.g. phenomenal content *is wholly a function of* [or *is wholly determined by*] representational content – or (what is the same) to the view that a mental entity has the phenomenal content it has *entirely in virtue of* its having the representational content that it has -- I am referring to the view that it is metaphysically, not merely causally, impossible that a difference in phenomenal content should fail to correspond to a difference in representational content.)

[306] In the present context, an 'experience' is an occurrent mental phenomenon having phenomenal content. So e.g. the occurrent phenomena posited by Chomsky and Marr and mediate visual and linguistic processing are not – as we are using this word – experiences (though, conceivably, others might describe them as such); and the non-occurrent mental contents – e.g. my belief that 2 is even – are not (as we are using the word) 'experiences'.

[307] In any case, this is how we will be using the word 'intentionalism.' In the literature, 'intentionalism' sometimes refers to the doctrine that I am calling 'representationalism', and sometimes it refers to the more restricted doctrine that the phenomenal supervenes on the representational.. In this paper, we will use that word in the latter sense only.

[308] Byrne [1].

[309] Byrne uses the word 'content' for what I am calling 'representational content', and he uses the term 'phenomenal character' for what I am calling 'phenomenal content.'

[310] It would, of course, be possible to perceive a blue chip without perceiving it *as* blue (consider e.g. color-blind people – who can see objects but can't see them as blue or red or green).

[311] Of course, the feature of B whereby it represents the redness of the car may not be a constituent in the *mereological* sense; it may not be a *part* of B in the sense in which a brick is a part of a building. Indeed, it almost certainly isn't such a part. It may be something much more abstract – some functional property, perhaps. So it may be a constituent of B only in a very abstract sense. But for our purposes it is enough that there be *some* mental property of B – *something* about B – whereby it represents the redness of the car. And this, no doubt, is the case.

[312] Henceforth, I will put points made by an imaginary objector in single-space.

[313] This counter-argument was suggested to me by a colleague who wishes to remain anonymous.

[314] Sometimes – indeed, usually -- the terms 'conceptual' and 'non-conceptual' are used instead. But I think that the distinction I have in mind is best understood as one between *propositional* and non-propositional contents, not between conceptual and non-conceptual content. For it is fairly clear – and, in any case, it is generally held -- that propositions can have non-conceptual content. The proposition *Socrates is bald* is *de re* about Socrates, and thus has Socrates himself – who is obviously not a concept – as a constituent. So the distinction between propositional and non-propositional content is different from that between conceptual and non-conceptual content – notwithstanding that some people talk of the distinction between 'conceptual and non-conceptual content' when they are discussing whether the content of a representation can be embodied in a *proposition.* Actually, the people who use the terms 'conceptual content' and 'non-conceptual content' to mark the distinction *I* am talking about – the distinction between, as I am putting it, propositional and non-propositional content – almost always hold that there *are* object-involving propositions, i.e. that non-concepts – e.g. Socrates, Plato – can be constituents of propositions. Given this, it is odd that, when they are discussing the question whether all represnetaitonal content can be embodied in *propositions*, they often state this question as 'is all representational content *conceptual*?' Again, it seems to me that (often, though probably not always) what is *really* being asked is 'is all representational content *propositional*'' (or 'is all represnetaitonal content capable of being embodied in a proposition?') and not 'is all represnetaitonal content *conceptual*?'

[315] What I mean by this can best be understood by contrasting visual representations with patently hybrid –patently heterogeneous -- representations. A comic strip is a patently hybrid kind of representation: there is a picture (e.g. of a superhero performing some great feat) and there are also words (which are put in a bubble emanating from the mouth of the featured character). So comics are both analogue and digital and are thus, in an obvious way, heterogeneous representations. Now when we consider *purely* analogue representations – e.g. visual perceptions – there seems to no such internal division.

[316] Now I must disclaim what I've said. First of all, it is extremely vague: I've totally failed to elucidate what I mean by 'packaging' – a notion which is obviously central the discussion in question. (But I'd like to think that what I mean is not totally obscure.) And also admit that I've dealt too briefly with the question whether all represnetaitonal content can be embodied in propositions. (Obviously MacDowell has a lot to say on this, and I haven't addressed what he says.) But, first, I think I've said *something* worth investigating; and second (to reiterate what I said before), it is pretty clear that perceptions that have representational overlap may have *no* phenomenal overlap: and this – I think it no overstatement to say -- by itself is a real stumbling black for intentionalism.

[317] It must be noted that Byrne makes it very clear that he wishes to defend the position that, given *any* two experiences – regardless of whether they are associated with the same sensory modality and, regardless of whether they belong to *any* sensory modality (e.g. itches and tickles, arguably, do not

belong to any of the five senses) – those two experiences cannot differ phenomenally without differing representationally. It is curious that Byrne holds this because the considerations *he* adduces support only the weaker thesis that given two experiences fro the *same* modality, they cannot differ phenomenally without differing representationally. (Recall Byrne's point about the blue chip, viz. you can't imagine a change in the phenomenal character of your experience of the blue without imagining a change in the color that the chip is represented to you as having. This is the pillar of Byrne's argument; but it obviously supports, at most, the claim that *within* a particular modality phenomenal variation supervenes on representational variation.

[318] There has been a great deal of discussion in the literature as to whether so-called 'spectrum-*inversion'* is possible. I do not know whether it is possible. But there seems to me nothing impossible about the spectrum-*shift* just described. It seems perfectly possibly – metaphysically, if not medically – to alter someone's brain in such a way that, in response to physical disturbance x1, they have phenomenal content y2, even though previously that person had phenomenal content y1 in response to x1. In fact, there seem to be cases where such spectrum-shift has *actually* happened. For example, the pianist Rosalyn Turek claims that the sensation which she associates with hearing (e.g.) a middle C has changed in the course of her life. (See *Reflections from the Keyboard* by David Dubal; Schirmer Books; 2nd edition; Nov. 1997.)

[319] I say 'at least arguably' because many will disagree with my contention here: they will say that, after the operation, p2 (for example) will *continue* to represent a C#; that, in general, for any n, p_n will *continue* to represent what it did before the operation. There is not really enough space for me to give this point of view the consideration that it deserves, but one criticism that can be made of it – and that surely cannot be summarily dismissed – is this. If that objection is right, then after the surgery, the person's perceptions of pitch are *systematically false*: p2 is associated with the representational content *middle C#*, when in fact the sound in question is a middle C; and so on *mutatis mutandis* for p3,p4 ...But the person in question could, post-operatively, recover his ability to make correct judgments about what pitch he was hearing. He could figure out that, when had *that* sort of phenomenal experience, the cause of it was a C#, not a C. He could, indeed, become every bit as adept at identifying what note was being struck as a Mozart. And, of course, he would make these identifications on the basis of his acoustical perceptions. So, given this, it seems arbitrary to maintain that those perceptions would all be delusive – i.e. that, as a result of the operation, he was simply *mis*-hearing pitches. It seems more natural to say he is perceiving pitches veridically, but that the way in which his phenomenal contents involved in these perceptions are paired off with their physical causes changed as a result of the operation. But, I grant, this is just an intuition of mine and will not meet with universal acceptance.

[320] 'SW' is short for 'Smith Earth.'

[321] 'TS' is short for 'Twin-Smith.'

[322] What a speaker means often diverges from what his words mean. If I tell an employee, 'I think it might be a good idea for you to clear your desk', I'm really telling him that he's been fired. But my words don't literally mean: you're fired. Their literal meaning is innocuous. See Chapter__for an extended discussion of this.

[323] See Burge (1979, 1982).

[324] 'BA' is short for 'Burge's anti-individualism.' In a moment, I'll define the term 'anti-individualism.'

[325] 'AU' is short for 'arthritis unpleasant.'

[326] See Burge (1979, 1982). Evans (1982) puts forth a similar argument.

[327] 'U1' is short for 'utterance #1.'

[328] 'MP' is short for 'Max plaid.'

[329] 'VR' is short for 'visual perception of robot.'

[330] 'CVR' is short for 'content of 'VR.'

[331] See also Russell (1948: 271-294).

[332] Gareth Evans (1982) was the first, as I know, to see that Burge's position has this consequence. But Evans, being a content-externalist, accepts this consequence, deeply bizarre though it is.

[333] 'MP' is short for 'Max plaid.'

[334] 'MP' is short for 'Max plaid.'

[335] G1 and G2 are supposed to be utterances of sentences, not sentences per se – sentence-tokens, not sentence-types.

[336] See Burge (1986).

[337] See Falvey and Owns (1994).

[338] See Kuczynski (2007) thorough discussion.

[339] "ARS" is short for "~~for~~ anything representationally symbolic."

[340] Hempel (1965).

[341] Hempel (1965).

[342] Quine (1977).

[343] Finally, others have tried to use theorems of (Bayesian) probability theory to solve this problem. (Bayes' Theorem: [P(A/B)=P(B/A)×P(A)]÷P(B).) (See Trout (2000).)

The term ‹P(x)› *is not defined* by the Bayesian. (The brackets are to be taken as quasi-quotation marks. Cf. the next footnote.) It has an assumed, but unidentified, meaning. So Bayesian probability-theory is an uninterpreted formal calculus; and, like any such calculus (setting aside those that are inconsistent), it admits of many different interpretations, at least some of which are unintended (meaning that, relative to those models, the term(s) that it was the purpose of that calculus to clarify are given meanings that are at odds with the intuitions by those terms, as ordinarily used, are, in effect, defined).

The term "probability theory," it must be pointed out, may refer to one of two things: the mathematical theory of probability, which aspires to identify the mathematical consequences of axioms in which expressions like "confirm" ("probabilify") and "probability" itself are undefined; and "probability theory" may also refer to the branch of philosophy that aspires to say what is meant by "confirm," "probability," and other expressions of their ilk. Carnap (1950) makes a similar point.

I won't revisit this point. Given the technicalities involved in doing so, the result would be word-inflation not counterbalanced by content-inflation. And this puzzle is inherently *non*-technical. Philosophical problems are *ipso facto* not formal problems. For, once a problem is formal, it becomes part of mathematical logic.

In any case, the Bayesian attempts to solve these problems all involve the same two mistakes. First, those theories take the term "probability" (and "confirms," etc.) as being undefined (or as being defined only in the sense in which the term "point" is defined in an uninterpreted system of geometry). From the viewpoint of the mathematical (and therefore the Bayesian) probability theorist, such terms are uninterpreted. But the very question raised by Hempel's paradox is: What is meant by "probability," "confirm," and other such terms? An uninterpreted formal theory is *ipso facto* incapable of answering

that question, since that question amounts to this: *Given* a model (that, if duly interpreted, might apprise us of the meanings of "confirm" etc.), what is the right interpretation of it?

Second, such theories take the *word-order* and, more generally, the strictly linguistic properties, of (i) and (ii) (or of whatever pair of sentences they're considering, when dealing with Hempel's paradox) as being material to the manner in which they are to be confirmed. But do "Brutus killed Caesar" and "Caesar was killed by Brutus" have different confirmational properties? Surely not.

[344]

After we put forth our own solution to Hempel's paradox, we'll examine those put forth by Hempel and Quine. But, since I don't want this paper *per se* (as opposed to this paper *plus* the footnotes, which *ex hypothesi* deal with (semi-)peripheral matters, our discussion of alternatives views of the raven-paradox will, like the material you are now reading, be footnoted.

In any case, most of what we'll say can be summed up in a few words. Their solutions are either counterintuitive or simply incoherent. Given what it is for sentences P and Q to be equivalent, it makes no sense, I submit, to say that P and Q be logically but not confirmationally equivalent. (‹P and Q are equivalent› [344] is true iff there is no world, metaphysically or epistemically possible, in which P is true and Q is false, or *vice versa*. If a given datum D confers more support on the one sentence than on the other, the one is *ipso facto* possible in some epistemically possible worlds in which the other is false.) But even if it can be shown that an internally consistent logic could be invented relative to which logical and confirmational equivalence pull apart, it seems excessive to throw out classical logic to solve this one problem. Suppose that, having been charged misrepresenting their financials to their shareholders, the owners of Enron said to the judge: "Your honor, we didn't misrepresent anything. Relative to the non-classical logic that we invented, our records are perfectly consistent with the facts." Not the best legal strategy. Hempel's solution---though ingenious, as we'll see, and though not involving *ad hoc* (meta)logical views---involves *acceptance* of the contention that red carpet fibers are no less, and no more, probabilistically relevant to "all ravens are black" than are black ravens (or non-black ravens). Even if Hempel is right, is it not worth trying to produce a solution to his paradox that is more congenial than his to commonsense, however myopic may sometimes be? In this paper, we'll see that Hempel's problem *can* (which is not to say "should") be solved in a manner congenial to commonsense-friendly and also with classical (meta)logical principles.

[345]

Bearing this in mind, let now (in Section 1.1) discuss an assumption we'll make concerning the semantics of categoricals.

The philosophers and logicians who write about (i) and (ii) (or other comparable sentence-pairs) make the following assumption as to what those sentences mean, to wit:

> (#) (i) is to the effect nothing is both a metal and also a thing that fails to expand when heated. (ii) is to the effect that nothing is both a thing that isn't expanding isn't also a thing that is made of metal and is being heated.

The people who write about Hempel's paradox tend to be logicians and philosophers of science who aren't interested in the semantics/pragmatics distinction and therefore, when talking about what a sentence S "means," don't say whether they're discussing, on the one hand, what S literally means or, on the other hand, what its occurrences convey. (And there is a big distinction to be made hear: what a sentence conveys is a function, not *just* of its literal meaning, but also of background understandings and other pragmatic factors. "I'm coming over to your house," may, when uttered, convey a threat (*you'll soon be dead, because I'm coming over to kill you*) or an assurance (*contrary to what you fear, you* won't *soon be dead, because I'll soon be at your house with your heart-medication*). At the level of literal meaning, such utterance is to the effect that x (the speaker) is going to place y (the addressee's home).

Bearing this in mind consider both

(TS) “triangles have three sides”

and

(i) “metal expands when heated.”

As they are ordinarily interpreted, those sentences are to the effect that, for reasons of logic (or mathematics) a thing can’t be a triangle without having three sides and that, for reasons of natural law, a thing can’t be metal if it doesn’t expand when heated.

Given only that that this is how occurrences of those sentences are typically understood, it does *not* follow that this is what such tokens literally mean.

My inclination is to think that, *at the level of semantics*, these sentences don’t have logical or nomic force, and it is through implicature, not explicature, that utterances of them transmit the logically/nomically pregnant claims just describes. I say this because those implicatures, if that’s what they are, can be cancelled and are therefore, by Grice’s (1975) plausible criterion, pragmatic epiphenomena (to use Kent Bach’s (1987) expression) as opposed to semantic explicatures. Cf.:

(iC[345]) “All metal objects expand when heated. But there’s nothing inherent in what it is to be a metal object that disposes such objects to expand when heated. Long ago, some alien prankster, hoping to mislead future scientists, destroyed all metal things that didn’t expand when heated.”[345]

Cf. also:

(TSC) “triangles have three sides. But triangles *could* have four-sides. This would happen if some triangular living creature grew a fourth side.”

(iC) is a highly fanciful claim, for which there is no evidence. And TSC is, to be blunt, a really stupid claim. But neither claim is tautologically false. In other words, it isn’t solely in virtue of facts about definitions and semantic rules (and the like) that either claim is false.

This is obvious in where (iC) is concerned. (iC) is empirical claim; it *could* be true. Therefore it isn’t tautologically (definitionally/semantically) false.

What about TSC? TSC is a statement that an actual student of mine---a superb writer, by the way---made in class. In fact, every semester, I get around 15 students who make statements not in any relevant way different from TSC, e.g. “1+1=3’ would be true if we found that two 1lb rocks jointly weighed 3lbs.” (In fact, this was Mill’s position, as well as the later Wittgenstein’s.) Most of these students have at least as good a command of the English language as I do. So it isn’t *linguistic* incompetence that leads them to utter the likes of TSC. It’s intellectual incompetence. Stupidity, in other words. (Sorry fellas!) Which means that TSC isn’t *linguistically* non-kosher. Which in turn means that it is *not* solely virtue of definitions (or semantic rules, etc.) that TSC false and, therefore, that it *is* in virtue, in part, of extra-linguistic truths. And these truths, clearly, are of a logical or otherwise non-empirical nature.

Here’s where this leaves us. (i) and TS, *as they* (or their tokens, rather) are customarily interpreted, don’t make purely extensional claims; they don’t *just* say that all the things in one category are also in some other category. As ordinarily understood, they say, not *just* that certain classes are sub-classes of others, but that there is some kind of principled basis for these facts about class-inclusion. But we’ve

seen that the intension-related propositions typically conveyed by (i) and TS are cancelable.

[345] Which suggests that, at the level of semantics, TS and (i) and their likes are to be interpreted, not as intensionally pregnant claims, but as affirming sterile, purely extensional ones.

Of course, the points just made by no means settle the delicate question of whether, in some contexts or some values of phi and psi, ‹all phi's and psi's› is, *at the level of literal meaning*, to the effect that a thing's a being phi is connected in some robust, not purely statistical way, to its being a psi. But, when taken in conjunction with the points we're about to make, they do suggest that (i) and (ii), if read as being such that tokens of the one *semantically* but not *pragmatically* parallel tokens of the other. Hempel's paradox is readily solved.

Or, to make a similar, but non-coincident, claims: Hempel's paradox is readily solved if we suppose that, although (i) and (ii) *do* coincide in respect of literal meaning, they have hidden (phonetically/orthographically unrealized) context-sensitive parameters, in virtue of which tokens of the one differ in respect of what they convey from tokens of the other. (In Russian, omission of a tense-marker has the same *semantic* function that, in English, is had by non-omission of such a marker. if one says "ya vam blagadarem"---"I grateful to you" (no (phonetically realized) tense-indicator)---one is, at the level of literal meaning, saying that one *is* grateful to one's interlocutor." It's not out of the question that, in English, there are phonetically unrealized context-sensitive parameters.

Supposing that, indeed, tokens of (i) *are* semantically parallel to tokens of (ii), my thesis is that Hempel's paradox can be solved in a way congenial to classical logic and commonsense, *the reason being that we're falsely assuming that tokens of (i) are tokens of (ii) in respect of what they affirm*, the consequence being that we wrench a fine system of logic and spurn one of our best (and, in a sense, our only) instrument for detecting truth (viz. our faculty of intuition), when what we should be doing is scratching the surface of language. So, as far as logicians and philosophers of science are concerned, there is no significant difference between the contention that

> ($) Hempel's paradox is to be solved by taking pragmatics into account,

and the contention that

> ($$) Hempel's paradox is be solved by taking hidden (semantic) context-sensitive parameters into account.

Of course, ($) and ($$) do (presumably) differ from a linguist's perspective. Even so, the present paper, if cogent, shows that *some* branch of linguistics can solve puzzles to which linguistics has always been assumed to be irrelevant.

[346] The "A" stands for "accidental."

[347] Ayer (1972).

[348] Hempel (1966).

[349] The "N" stands for "nomic."

[350] The "W" stands for "wide-scope," the idea being that the nomic operator is given maximally wide-scope, since the whole of "for any x, if x is metal, then expands" is governed by it.

[351] **4.0 Hume-internal solutions to the raven paradox: Quine's solution**

Quine's (1977) solution is to say that that logical equivalence doesn't guarantee confirmational

equivalence. This idea is totally incoherent. P and Q are equivalent if they are true in exactly the same worlds. Thus, supposing that P and Q are equivalent, it immediately follows that, given any proposition R, the likelihood of R's being true in a world where P is true is identical with the likelihood of R's being true in a world where Q is true. Quine's proposed solution is thus a non-starter.

4.1 Hume-internal solutions to the raven paradox: Hempel's solution

Hempel proposes a more plausible solution. It obviously *seems* as though the existence of a pink piano provides less support for:

(1) "All ravens are black"

than does the existence of a black raven. And it obviously *seems* as though the existence of pink pianos provides more support for:

(2) "All non-black things are non-ravens"

than it does for (1). But, says Hempel, these intuitions are misleading. Under ordinary circumstances, if we know that something is a piano, we almost certainly *already* know that it isn't a raven.

But imagine the following. You have bet your friend Smith $1,000,000 that all ravens are black. There is some object x such that, although you and Smith don't yet know it, x is a piano. All you know about x is that it's pink. In that case, when you *discover* that x is a piano, and thus a non-raven, you (but not Smith) will breathe a sigh of relief; and your confidence in (1) will surge, and so, of course, will your confidence in (2).

Now imagine a different scenario. You have bet your friend Smith $1,000,000 that all non-black things are non-ravens. There is some object y such that, although you and Smith don't yet know it, y is black. All you know about y is that it's a raven. When you discover that y is black, you will breathe a sigh of relief; and your confidence in (2) will surge, and so, of course, will your confidence in (1).

Thus, to Hempel, the confirmational disparities between (1) and (2) are illusory. When we imagine ourselves trying to confirm (2), we assume that we are looking at things that we *already* know to be non-ravens, but we don't make the same assumption *mutatis mutandis* when we imagine ourselves to be trying to confirm (1).

Hempel's solution is better than Quine's, but it's still no good. The problem is that, if it's known that there is no nomic or causal relationship between a thing's being a raven, on the one hand, and its being black, on the other, then there is *ipso facto* no reason to believe that the next raven you see will be black. Maybe you've seen a million ravens, all of them black. It's doesn't matter. Unless you have reason to believe that a thing's being a raven is connected somehow to its being black, you're guilty of the gambler's fallacy if, on the basis of that data, you believe that the next raven you see is (likely to be) black. But if you *do* believe that a thing's being a raven is connected in some way to its being black, then your confidence in (1) is derivative of your confidence in:

(1*) a thing's being a raven provides a basis of some kind for its being black.

We've already seen why. And we've already see why (1*) isn't equivalent with:

(2*) a thing's being non-black provides some kind of a basis for its being a non-raven.

If the existence of a black raven is to provide any legitimate support for (1), there must be a causal or nomic connection between ravenhood and blackness. Once this is realized, two things immediately

become obvious. First, Hempel's solution is spurious, since it presupposes that black ravens can confirm (1) *without* there being any connection of the sort just described. Second, (1) must be read as having nomic force. Since, thus read, it is no longer equivalent with (2), the paradox vanishes.

[352] Republished in Quine (1966).

[353] It is sometimes propositions that are described as "formally true," but this is quite rare, and it's almost always expressions that are so described. But, even though the terms "formally true," "formally entails," etc. are typically used in connection with sentences, there are, I believe, propositional analogues of the concepts they express. Consider the *proposition* (not any one of the corresponding sentences/sentence-tokens):

(i) 1+1=2.

(i) entails each of:

(ii) for some x, 1+x=2,

and

(iii) there is some x such x is even and x is 1's successor

and, finally,

(iv) there are continuous functions that cannot be differentiated at any point.

It's only in a thoroughly sterile, strictly model-theoretic sense that (i) entails (iv). The way in which (i) entails (ii) isn't so antiseptic, and neither is the way in which it entails (iii). But, it seems to me, the way in which (ii) follows from (i) is more transparent than the way in which (iii) does so; and a viable theory of propositions ought to validate that point. In my book__, and my paper__, I argue that, whereas first entailment holds in virtue of the structures of the *immediate* constitutents of the propositions involved, the second entailment holds in virtue of structures of those constituents---it holds in virtue of, as it were, the constituencies of those propositions' constituencies.

Moving onto another point: It is, I believe, sentence-*tokens*, never sentence-types, that are true or false. My reasons for this, which I state at length in my book__, are extensions of arguments put forth by Strawson (1950), Barwise and Perry (1983), and Kaplan (1989). Long story short: It's incontrovertible that no sentence-type containing a context-sensitive parameter is true or false. Consider the sentence-type "that guy is a threat to our way of life." Some *instances* of that sentence-type are true or false, but the sentence-type *per se* is neither. *All* sentences of natural language contain context-sensitive component. In some cases, e.g. "there are infinitely prime numbers," that component (in this case, the present-tense marker) aren't doing any work. But that has to do with the non-semantic fact that mathematical reality doesn't change. If, as Mill and other hyper-empiricists hold, it *does* change, then whether "there are infinitely primes" is true or false is a function of the time (and possibly the place) of utterance. Sentence-tokens are spatiotemporal objects. Sentence-types, being things of which there are instances, are properties and therefore aren't spatiotemporal objects. So, where natural language is concerned, it is always spatiotemporal entities that have truth-values. Of course, there could exist, and presumably *do* exist, context-*in*sensitive languages (cf. the artificial language, or language-extension, to which "1+2=3" belongs). But this would not be a good move. No difference is greater than that between the spatiotemporal and the Platonic. Given this fact, along with the fact that the term "truth-evaluable expression" corresponds to an explanatorily fecund category, it's hard to see how that

class could contain both expression-tokens *and* expression-types. For an extensive defense of this position, see___.

[354] Some would hold that *any* given sentence of natural language *ipso facto* fails to be *formally* true and that, so far as one isn't speaking false in describing (i)-(iv) as "formally true," one is saying elliptically that, given their syntactic structures, it's obvious how to associate them with sentences, belonging to artificial languages, which, unlike them, really *are* formally true.

Supposing, if only for argument's sake, that this is correct, it has no bearing on the analysis which we'll be defending. For that analysis, as we'll see, is no less consistent with the assumption "if snow is white, then either snow is white" is formally true than it is with the corresponding assumption concerning their artificial counterparts.

[355] More precisely, (#) and (##), taken jointly, satisfy two conditions. First, there is a one-one correspondence between the one's (proper) constituents and the other's. (Actually there are 9 such correspondences; but, of course, the relevant one, which is the one we'll discuss, is the one that pairs off "Bill" with "Laura," "hates" with loves," etc.) Second, supposing that P# and P## are the propositions borne by (#) and (##), respectively, the *relation* borne by the meaning of "hates" to P# coincides with the *relation* borne by the meaning of "loves" to P##, the same thing *mutatis mutandis* being true of each of the other two constituents of (#).

The reason that (#) does *not* have the same syntax as:

> (###) "either Igor went swimming or the U.S. economy is probably going to continue to deteriorate"

is that, unless (###) is parsed in a patently artificial manner, its parts cannot be put into a correspondence with those of (#) and *a fortiori* there is no correspondence R such that, relative to R, the relation borne by any given (proper) constituent of the one sentence to the proposition thereby meant coincides with the relation borne by that constituent's R-counterpart to the proposition meant by *its* host-sentence.

Syntax, in the linguist's sense of the word, is semantic structure. Two sentences are syntactically identical (or isomorphic) iff the semantic *relation* borne by a given constituent of the one sentence to the meaning of that sentence as a whole coincides with the semantic *relation* borne by that constituent's counterpart in the other sentence. So it's true---indeed, it's a truism---that syntax is expression-form. But it's a blunder to conclude from this old saw, as so many do, that syntax has nothing to do with semantics. Syntax is semantic structure. So one is contradicting oneself if one argues, as Fodor (1987, 1987b, 1990, 1994) and Davidson do, that a semantics-insensitive entity (e.g. a Turing-machine) can be syntax-sensitive.

Also, syntax is but *one* kind of expression form, as Strawson (1952) long ago made clear. For there is at least arguably *a* sense in which "Brutus killed Caesar" and "Caesar was killed by Brutus" have the same form---they have the same logical form, presumably, given the one sentence's inferential liaisons---what, and by way of what intermediate steps, it entails and is entailed by---coincide with the other's (with some minor qualifications, e.g. that "Caesar was killed by Brutus" follows more directly from "Caesar was killed by Brutus" than does "Brutus killed Caesar.") Thus, it isn't because of any structural differences between them that they have different meanings.

[356] In the opening paragraph of the *Principles of Mathematics*, Russell (1903) puts forth an analysis very similar to Quine's.

[357] Quine (1970).

[358] Frege (1884).

[359] These brackets are known as “quasi-quotation marks.” They are like quotation-marks *except* that they don’t quote the free variables that they enclose. What is the purpose of quasi-quotation marks? Consider the expression: “if *P* and *Q* and *R*, then *R*.” That expression isn’t true or false. So the sentence “For any propositions *P*, *Q*, and *R*, the sentence ‘if *P* and *Q* and *R*, then *R*’ is true” is *not* true. But the sentence “for any propositions *P*, *Q*, and *R*, a true sentence results when, given the open-sentence “if *P* and *Q* and *R*, then *R*,” the variables are uniformly replaced with arbitrary sentences” is true. And (A) is a concise way of saying this. In general, quasi-quotation marks make it possible to express generalizations about sentences in a concise way. Quine (1940) was the first to see the need for quasi-quotation marks and duly introduced them in his (1940) book *Mathematical Logic.*

[360] Wittgenstein (1922).

[361] *Some* of its instances are formally true, e.g. “either Jim is tall or Jim is not tall or Jim is tall or Jim is not tall.”

[362] Short for “all truths.”

[363] Short for “it’s raining.”

[364] Why sometimes? Because one must have an assurance that one is dealing with sentence-tokens whose actual structures aren’t hidden or distorted by their surface structures. One can infer “Bill is a duck” from “Bill is a big duck,” but not from “Bill is a decoy duck.” One can infer “Bill is good” from “Bill is a good person” but not from “Bill is a good torturer.” When dealing with sentences of English, or of any other natural language, one has to know what a sentence means *in order* to know to what degree its surface structure is representative of its logical form. So, where sentences of natural language are concerned, judgments about form must be checked against knowledge semantics. Which means that, when given a sentence belonging to a natural language, any legitimate judgment concerning its truth-value or inferential liaisons must be semantics based. Which in turn means that form-driven judgments are legitimate only if its assumed that the surface-structures of the sentences in question do justice to their actual structures. And it’s only in very specific, and very artificial contexts that one has any such assurance.

[365] In what remains of this chapter, all of the indented expressions are sentences or open-sentences. For that reason, it isn’t necessary to enclose an expression in quotation-marks to identify it as an expression that picks out a sentence, as opposed to a proposition; and I’ve thus not enclosed the indented expressions in quotation-marks.

[366] In this chapter, there are many places where, technically, I should have used quasi-quotation marks but where I instead used ordinary quotation marks or didn’t use any quotation marks at all. I didn’t want an excess of obscure symbolism to obscure matters of substance.

[367] “LC” stands for “logical constant.”

[368] The viewpoint I am advocating, though by no means particularly non-obvious and by no means heterodox (relative to certain a certain tradition of thought, beginning with Leibniz (cf. the *Leibniz/Clarke Debates*) and including Einstein) isn’t even in considered (so far as I can tell) in the contemporary leading works on this topic. See Gale (1968), Gribbin (1985), Pagels (1985), Lewis (1986), Nahin (1997, 2001), Miller (2005). Pickover (1999), Davies (1996, 2002), Randles (2005)

Toomey (2007). That said, that viewpoint is anticipated, though not explicitly endorsed, by Poincaré (1913), Einstein and Infeld (1959), Einstein (1962), Russell (1928, 1948), and Reichenbach (1956, 1958). I was, frankly, a bit surprised that Lewis didn't mention it, given his high level of interest in the subject and his unparalleled level of logico-philosophical acumen. Whatever demerits might afflict my contention and my argument for it, it seems sufficiently consonant with intuition that it should be considered, if only to be rejected. But much of what the aforementioned contemporaries say simply assumes its falsity, much to the detriment of their argumentation, making it hard to believe that they've given this viewpoint their due. I apologize in advance to any author who *has* made points identical to or coincident with mine and welcome a dialogue with her or him.

[369] Also, any given spatiotemporal thing is moving relative to something.

[370] Obviously (i) isn't *formally* false. It isn't like "snow is white and it is not the case that snow is white." Nor is (i) tautologously false. It isn't like "there are not three feet in a yard." (i) is *analytically* false. It embodies an incoherent assumption about the nature of persistence in time.

[371] To use Fodor's apt expression. See his paper "The Present Status of the Innateness Controversy," in Fodor (1981).

[372] Kant makes a similar point at the very beginning of the *Critique of Pure Reason.*

[373] I later found that his findings coincided with some of mine (cf. Chapter 21 of *Mind, Meaning, and Scientific Explanation,* Dubuque 2010, where I argued, in an attempt to legitimate psychodynamic psychology, for a specialized version of Humphries' thesis.) That said, for the first 90 minutes of his talk, I fought his contention tooth and nail.

[374] I do not believe that this is all that is involved; indeed, I myself hold that our knowledge of laws are based on our knowledge of mini-explanations—we know about the inverse-square law only because we *already* know that my releasing the pencil caused it to drop.

[375] Which is one reason why Russell's (1903) of propositions---propositions are the situations in virtue of which they're true, when they're true---is false, it also being why Frege's dictum that "a fact is a true proposiistion" is false, given that true propositions cannot possibly be non-platonic objects, given that false propositions are platonic objects. This also completely invalidates Nathan Salmon's (2007) theory, which he has urged for 30 years, that Socrates *per se* is a constituent of the proposition *Socrates was wise.* Salmon grants that Socrates doesn't exist, and he also grants that his existence is necessary for that of that proposition. He 'solves' this dilemma by saying that non-existent propositions exist.

[376] On the basis of knee-jerk interpretations of Kripke's work, Salmon disagrees. His viewpoint---see below---requires him to hold that it's rational to believe in square circles and, moreover, that the fountain of youth cannot possibly fail to exist.

[377] They're truths of a non-conceptual, non-propositional kind; but that subtltety can and should be ignored here, since what matters is that data, unlike non-representations (e.g. rocks), are things to which the concepts of accuracy (truthfulness) apply, and data are therefore, in the present context, relevantly proposition-like.

[378] See Chapter 21 of Kuczynski (2010).

[379] I am not alone in this. Kripke, Ernest Nagel, and Einstein himself all reject it.

[380] Davidson 1984.

[381] Carnap (1947) makes this point.

[382] See Davidson 1984, Evnine.

[383] Russell (1903) said that "P is a proposition" is equivalent to "P entails P". He is right. But that probably doesn't qualify as an *analysis* of the concept of a proposition. (I don't think Russell was trying to give an analysis.)

[384] Sentence-types – the types of which those tokens are instances -- seldom if ever encode propositions. The reasons for this don't concern us -- even though those reasons have led, by way of totally spurious reasoning, to hyper-contextualist and nihilistic views about meaning that we won't go into and that, I think, you wouldn't find plausible anyway.

[385] Nonetheless, not everything in a sentence (or, sentence-token, rather) expressing a proposition *denotes* a property. Some sentential material denotes nothing, and merely encodes material – usually identifiable with so-called "grammatical inflections" -- saying how to identify a third property on the basis of two previously identified properties. This is further discussed in Conceit # 8 and extensively further discussed in the rest of the paper. Basically, grammatical inflections encode functions that assign properties to n-tuples of properties; and it is these n-tuples that distinguish "Aristotle strikes Socrates" from the meaningless heap "Aristotle, the relation of striking, Plato." So even though everything in a proposition is a property, not all the information in the sentence that *expresses* a proposition a property.

[386] In a number of places, Jerry Fodor rightly emphasizes that any theory of propositional content must satisfy this desideratum.

[387] This is probably the most counterintuitive aspect of my analysis. But we will find a great deal of justification for it.

[388] Frege talks about functions, not properties, be saturated. So he would speak of the one-place function *x snores* being "saturated" by Plato. In this context, the difference, if there is any (which I don't believe there is) between properties and functions is irrelevant.

[389] Though there will be some dispute as to whether possession supervenes on other, non-object-involving properties.

[390] For expository simplicity, we're leaving aside niceties relating to tense-markers and place-indicators.

[391] At least if Frege is right.

[392] It would actually be better to say: "while C and C are both Tommy-like", since this expression, unlike the expression "since C and D both comprise Tommy," avoids the false metaphysical

implication that Tommy is a veritable, detachable constituent of the situations in which he is embedded.

[393] My own view is that Hesperus-Phosphorous problems are dealt with by taking into the enormous and, by our semanticists, always overlooked information involved in grasping singular propositions. Once that pre-semantic data, as I call it, is taken into account, there no difficulties in the slightest in accounting for the fact that there can, under many circumstances, be gross differences in respect of cognitive value between "Hesperus is lovely" and "Phosphorous is lovely." So I will take leave of this particular concern.

[394] or, in any case, from post-Fregean such as Montague

[395] This is, of course, not the only analysis of *nothing snores.* We can also identify that proposition with a set S whose members are the following:
a The property of being universally uninstantiated.
b. The property being identical with the property of snoring.
c. The property of being identical with the universally uninstantiated property of snoring.

First of all, notice that c is derived by Conceit# 8 from a and b. Any instance of c is an instance of (b) the property of being identical with the property (which is *not* the same as being an instance of snoring) and is also an instance of (a) the property of being universally uninstantiated. So if c is instantiated, so are b and a. Thus, if c is instantiated, then there is an instance of the property of being identical with a universally uninstantiated property of snoring. In other words, there exists a universally uninstantiated property and that property is identical with the property of snoring. Since c cannot be instantiated without a and b being instantiated, that means that if a-c are instantiated, then there exists a universally uninstantiated property, that property being the property of snoring. And that, of course, is equivalent (if it isn't outright said) by "nothing snores."

[396] I know that this point was made by John Perry, but I have forgotten where he says it and haven't been able to find this out.

[397] Different people have different views as to what that property. My (Kripke-influenced) guess is that it is the property of developing into something with certain biological and, hopefully, psychologically characteristics out of a certain gamete. But whatever that *property* is, it *exists* even when Joey doesn't, i.e. even when it isn't instantiated.

[398] It was Donald Smith who brought it to my attention that my analysis avoids this pitfall of direct reference theory. I am sympathetic to many aspects of direct reference theory. I think that there is some x such that, if true, the sentence "Socrates is bald" is true exactly if the same is true of the singular proposition *x is bald*. So I hold that, for that sentence to be true, it is necessary that Socrates – not *the great hemlock-drinking philosophers* and not *the sarcastic main protagonist in most of Plato's dialogues* – be bald. Basically, I was, and am, a card-carrying direct-reference theorist. But, independently of that, I held the view, described in this paper, about propositions. It was not until talking with Donald, some weeks ago, that I could use theory of propositions to produce a form of direct-reference theory wouldn't be stuck with the embarrassing (because so implausible) consequence that the meaning of "Joey likes to read" vanishes when Joey does – a consequence that Soames and Salmon deal with by saying that "Joey likes to read" does indeed encode a proposition: a non-existence one. (So, in their view, there exists a proposition that is the meaning of "Joey likes to read" that doesn't exist.)

[399] My own view is, in a way, not far from Frege's. For I hold that, indeed, it isn't Socrates *per se*,

but is rather the property of being identical with Socrates (which, of course, is uniquely satisfied by Socrates, so far as anything satisfies it) that is a constituent of the proposition *Socrates is bald.* Frege held that the relevant property/concept was a different one – one along the lines of great hemlock-drinking Greek philosopher. But, in this context, this difference doesn't matter terribly.

[400] Here's a more graphic way of putting it. We want to find out if there could possibly be square circles. Lewis rightly says that we can figure this out, and that our doing so doesn't involve leaving our own world. All we have to do is manipulate the materials – the properties, presumably – that are available to us right here: the properties *square*, *circle*, and maybe a few others. Having done the combinatorics, it's clear that, no, there can't be square circles. But this suggests that it is because is defectively put together, not because of what is going on (or not) in other worlds, that *there are square circles* is incapable of being true. What is
going on in other worlds, if anything, is, at best, a reflection of these inter-property relations.

[401] At one point in *On the Plurality of Worlds,* Lewis says that we shouldn't be so sure that those propositions are distinct, given how "theoretical" the notion of a proposition is. This is obviously desperate thinking. Whatever they are propositions are the things we affirm, believe, doubt, etc. And somebody can believe that 1+1+2 without believing that triangles have three sides.

[402] Even Quine gave up, conceding that sets were incorrigibly platonic and that sets were ineliminable.

[403] Although the word 'concept' is ambiguous, that ambiguity is not as radical as that associated with words like 'bank' and 'rent'. Presumably, a mental entity qualifies as a concept_O in the psychological sense – as a *grasp* of something – if it somehow encodes a concept_O in the *objective* sense. I grasp the concept_O of triangularity – which is a non-mental entity – by virtue of having a mental content which somehow encodes that concept_O. So a concept in the psychological sense is a mental entity which encodes a concept_O in the objective sense.

Also both concepts and conceptso are *modes* of presentation. The concept_O *class of all points equidistant from a given point* is a mode of presentation of the property of sphericity, and so is the concept_O *closed three-dimensional figure of uniform curvature.* So underneath the superficial ambiguity of the word "concept", there are identities of meaning.

[404] Of course, you could think about your grasp of Socrates. But what enables you to think about Socrates is that you *have* a concept of Socrates, not that you *grasp* your concept of Socrates. You don't have to grasp your own concept_O – your own grasp -- of Socrates to think about Socrates.

[405] I think that intentionalists might say that the very ability to have the phenomenology involves a background of representational abilities.

[406] Fodor (1998:13).

[407] Fodor (1998:35).

[408] Fodor (1998: 80-86).

[409] Fodor 1981e.

[410] Fodor (1981e:283-292).

[411] Fodor (1981e:283-292).

[412] Fodor (1998:46-47).

[413] See Kuczynski 2004b, where the following argument is given in much more detail.

[414] The reason is that we give "wide scope" to the descriptive information. I tell you "some guy stole my wallet." You then think to yourself: *he* didn't have to become a thief – *he* might have become a police officer. Here you are *not* thinking: *it is possible that somebody x uniquely stole __'s wallet and also that x did not become a thief.* What you are thinking is: *somebody x uniquely stole __'s wallet and there is a possible circumstance where that x did not become a thief.* So the descriptive information is given *wide* scope with respect to the modal operator. Even though it is always through descriptive information that we acquire concepts of objects (or so I will argue), that doesn't corrupt our modal intuitions; for we give the descriptive information "wide scope" with modal operators.

[415] Or *seemed* to argue: see below

[416] Jerry Fodor (1998) has gone so far as to say that a causal connection to x is *all* that is involved in having a concept of x. So I can have *no* beliefs about the number and yet have a conception of it: I need only have a casual connection to it (or, more accurately, some instance of it). I need only have a causal connection to Socrates to be able to think about him. I needn't know *anything* about him. My book__ is a discussion of Fodor's atomism and his arguments for it.

[417] It is possible that what Kripke says is not so much that people know of *no* individuating descriptions of the things they think about and refer to. His point is that the individuating descriptions are not the ones that Russell (and possibly Frege) *thought* they were. I think that Kripke himself says something similar to what we say below – when we argue that one *does* need an individuating descriptions *of a very special kind* to refer to an object. (I won't press these exegetical points. I will argue only for the point that none of Kripke's cogent points to any degree warrant the rejection of the "descriptivism" I have been urging.)

[418] What I will say goes a bit beyond what Kripke says; my own commentary is interlaced with an exposition of Kripke's views.

[419] More exactly, that person gives Socrates the Ancient Greek equivalent of "Socrates", whatever that is. In this discussion, we will ignore the fact that, over the years, names may undergo considerable phonemic change. Obviously "Confucius" sounds very little like the name of the old sage. These phonemic changes introduce no philosophical mysteries, but allowing for them would clutter up the exposition considerably.

[420] In fact, it is arguable that *all* such descriptions do, except when the description in question refers to some abstract object, like the number two. The (often hidden) presence of indexicals is needed to prevent problems relating to what Strawson called the "massive reduplication". Your friend Fred says to you "somebody uniquely invented bifocals; his name is 'Ben Franklin'". As it happens, on a distant planet, some other person (or intelligent creature) also invented bifocals. Is Fred *wrong*? Not if we take him to mean: in *this* context – in the context of the world *we* occupy – somebody uniquely invented bifocals. Surely we would not take Fred to rule out the possibility that on some distant planet there were intelligent, bifocal-inventing life-forms. So, in effect, we attribute a hidden indexical to Fred's statement. For reasons of this kind, we must see our discourse as permeated by indexicality, hidden though it may often be.

[421] Of course, the term “occasion” has to be defined somewhat narrowly. One can refer to twenty different people in a single conversation and thus on a single “occasion”. But at any *specific* juncture of that occasion – at any *one* time (leaving aside ambiguity) – one is referring to just one person or object. When I say “occasion” I mean a specific juncture of a discourse.

[422] See Russell 1948 (89-99).

[423] Russell may well have chosen historically pregnant examples for purely illustrative reasons.

[424] It isn’t even clear if Russell *was* wrong about the nature of the uniquely individuating descriptions. It could be that he was simply *using* historically pregnant descriptions to illustrate his basic point, which is that one must have *some* uniquely individuating description of Socrates to think about, or refer to, Socrates. If he had illustrated his point with some uniquely individuating description that didn’t *obviously* apply to Socrates – e.g. *the guy to whom Seymour was referring on occasion C…*--nobody would have followed what he was saying. So his use of uniquely individuating descriptions like *the philosopher who drank Hemlock* may not really indicate his belief that *that* is an actual uniquely individuating description through which one might refer to, or think about Socrates. It might just be a superficial expository device. But I am not a Russell scholar; so I remain open on this.

[425] I say “may” for a couple of reasons. After one learns that (so to speak) Hesperus=Phosphorous, then *Hesperus has phi* may end up coinciding in cognitive value with *Phosphorous has phi*. (Of course, this point has to be adjusted for speakers of other languages.)

Also, depending on how one initially learns the meaning of “Hesperus” and “Phosphorous”, *Hesperus has phi* *Phosphorous has phi* might never have had different cognitive values (for that person) to begin with. This will be the case if am introduced to the words by somebody who points to Venus and says: “*that* is called both ‘Hesperus’ and ‘Phosphorous’”.

[426] Let me end by considering an objection to this last line of thought (I will put the objection in the mouth of an imaginary locator):

“You have *ex nihilo* produced this strange theory of conception as being sustained by knowledge of interlocking existence claims. And you claim that recognition consists in seeing that some one object satisfies two existence claims. But all of this seems very *ad hoc* – something you’ve confabulated to get your bankrupt theory off the hook.”

I think that this comment has force only to the extent that we’ve been jaded by wrong theories of sense-perception and, more generally, of objectual awareness.

When you see Bob, embedded in your sense-perception is existential information. You cannot get rid of that existential information without eliminating the perception. This suggests that that existential information is constitutive of the content of the perception. Granted, Bob himself is (part of) the representational content of that perception. But he is given to you via this existential information.

So when you see Bob on some occasion t, the content of that perception is given by some bit of existential information like: *there is some chap whom I’m now seeing, wearing a bowler hat…*

You see Bob again, on occasion t*. Again, the content of your perception is given by some existence claim: *there is some chap I’m seeing, wearing a baseball cap…*

On both occasions, the content of your perception is existential.

Suppose that, on t*, you *recognize* the person you are seeing is the person you saw on t. Your t-perception is given by existential information, and so is your t*-perception. So if, on t*, you *recognize* that you are seeing the same person you saw on t, that must consist in your realizing that the same entity satisfies both existence-claims are satisfied – that there is some one entity who satisfies both. Once a few platitudes are granted about sense-perception – that sense-perceptions have existential

content -- our theory of conception and recognition proves to be far from *ad hoc*.

[427] The tie-shop example, and many of the lessons I extract from it in the following pages, are of course taken from Sellars (1963a: especially 144-149). Some of what I say goes beyond what Sellars (1963a) says or, perhaps, even implies. But I would like to think that what I say here does develop some of the inner meanings of that great work.

[428] The property of being perceivable is a matter of degree, and also of context. Therefore, the same is true of the property of being theoretical. What you can *see* is a function, at least in part, of what your background theories are. (This is a very Sellarsian point.) But, rough and ready though it is, the concept of a perceivable property is not without its utility; and the fact that it is subject to qualifications does not have any bearing in this context.

[429] Burge 1993.

[430] I realize I am splitting an infinitive. But the context licenses a departure from the rule (questioned my many a linguist) that infinites should not be split.

[431] Sellars (1963b: 336).

[432] Byrne (2001) makes a comment very similar to this to argue against some alleged counter-examples to intentionalism due to Peacocke. Byrne's use of this kind of attack works against *Peacocke's* counterexamples, but not against mine. There is little resemblance between my counterexamples and Peacocke's.

[433] Byrne (2001).

[434] Of course intentionalists might say: ' no one is mad enough to suppose that *different species* have the same kind of phenomenal content in response to a different kind of sound (experienced under a given set of circumstances). And, of course, no one denies that creatures that have different phenomenal contents (in response to a given kind of sound under a given set of circumstances) can hear just as veridically, if not more so, than we. Intentionalism – at least any sane version of it – applies *within* species, not *between* species.'

But this concession eviscerates intentionalism. Intentionalism wants a metaphysically necessary relationship between the representational and the phenomenal; not a merely causal, contingent one. Obviously if intentionalism is not true *across* species, then a given bit of representational content *can* be realized by very different phenomenal contents, and there is no long any kind of reduction of the phenomenal to the mental, and thus no *explanation* of the phenomenal in terms of the representational (at least none along intentionalist lines). But the interest of intentionalism is surely not that it register the idiosyncrasies of species, but that it captures some kind necessary fact about the relationship between phenomenal and representational content. And if this relationship doesn't hold across species, then that relationship isn't necessary. And it probably – almost assuredly - *doesn't* hold across species.

[435] The asterices are meant to be quasi-quotes.

[JK1] I added a footnote here.
[JK2] I changed the wording a bit here.
[JK3] I added this parenthetical comment

[JK4]I added the word ‘qualitatively’
[JK5]Again, I added the word ‘qualitatively.’
[JK6]This is probably the most natural place to divide the paper. I would be happy to supply any connecting passages, if you think it appropriate.
[JK7]I added a paragraph break and replaced a long sentence with a shorter one.
[JK8]I added the missing data here (concerning page numbers and publication places)
[JK9]